

# BIOMETRIE UND INFORMATIK

## IN MEDIZIN UND BIOLOGIE

bisher  
EDV in Medizin und Biologie

Abt. Medizinische Informatik und Statistik

Gustav Fischer Verlag Stuttgart  
Verlag Eugen Ulmer Stuttgart

**Band 20**  
**Heft 1-2/1989**  
ISSN 0300-8282

# BIOMETRIE UND INFORMATIK

IN MEDIZIN UND BIOLOGIE

bisher EDV in Medizin und Biologie

**Schriftleitung**  
**Prof. Dr. Hans Geidel, Stuttgart**  
**Privatdozent Dr. Walter Lehmacher, München**  
**Dipl. Math. Joachim Vollmar, Mannheim**  
**Gustav Fischer Verlag Stuttgart**  
**Verlag Eugen Ulmer Stuttgart**

**Herausgeber**

P. Bauer (Köln)  
H. Geidel (Stuttgart)  
G. Grabner (Wien)  
W. Lehmacher (München)  
H. Rundfeldt (Hannover)  
W. Stucky (Karlsruhe)  
H. Thöni (Hohenheim)  
J. Vollmar (Mannheim)

## Inhaltsverzeichnis

Vorwort		1
Editorial		2
Statistische Expertensysteme	R. Haux	3
An exact nonparametric randomization test for censored data	R. A. Hilgers N. Neumann	66
Besondere Anlagentypen und ihre varianzanalytische Auswertung	E. Hofmann M. Precht	74
Buchbesprechungen		86
Persönliches		87

### Redaktion:

**Schriftleiter** (verantwortlich im Sinne des Presserechts): Prof. Dr. H. Geidel, c/o Verlag Eugen Ulmer, Postfach 70 05 61, 7000 Stuttgart 70, Tel. 47 28 80.

**Verlag:** Vertrieb und Werbung: Gustav Fischer Verlag GmbH & Co KG, 7000 Stuttgart 70, Wollgrasweg 49, Tel. (07 11) 45 80 30, Postgirokonto Stuttgart 13 556, Stuttgarter Bank 45 290. Herstellung und Anzeigen: Verlag Eugen Ulmer GmbH & Co., Wollgrasweg 41, 7000 Stuttgart 70, Tel. (07 11) 45 07-0, Telex 7-23 634, Postscheckkonto Stuttgart 74 63-700, Zürich 80-47072, Wien 1083.662. Deutsche Bank AG, Stuttgart, Kto. 14/76 878, Südwestbank AG, Stuttgart, Kto. 21 000. Herstellung Siegfried Märkisch. Verantwortlich für den Anzeigenteil: Dieter Boger, z. Z. ist die Anzeigenpreisliste Nr. 7 gültig. Anzeigenschluß: Am 10. der Erscheinungsmonate Februar, Mai, August, November

**Druck:** Kaiser-Druck GmbH, 7335 Salach, Schulstraße 11, Telefon (071 62) 70 11 + 70 12.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder ähnlichem Wege, bleiben vorbehalten. Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Biometrie und Informatik in Medizin und Biologie ist eine deutsch/englischsprachige Zeitschrift, die theoretische und anwendungsbezogene Arbeiten aus dem Gesamtgebiet der Biometrie und Informatik in Medizin und Biologie enthält. Die Zeitschrift veröffentlicht sowohl Beiträge über neu entwickelte Ansätze als auch über neue Anwendungsmöglichkeiten bekannter Methoden. Zusätzlich soll dem tutoriellen Aspekt durch Übersichtsartikel Rechnung getragen werden. Der Bezug zur Medizin oder Biologie darf dabei jedoch nicht vernachlässigt werden. Jedes Heft enthält neben einem Leitartikel u. a.

- Übersichten zu bestimmten Themenkomplexen
- Originalien
- Informationen über Programme und Programmsysteme
- Aktuelle Informationen zu Veranstaltungen und Neuerscheinungen auf dem Buchmarkt.

Die Zeitschrift wird zur Zeit referiert in den „Biological Abstracts“ und „Current Index to Statistics“.

Biometrie und Informatik in Medizin und Biologie erscheint vierteljährlich. Der Bezugspreis beträgt jährlich DM 248,- (Einzelheft DM 75,-) zuzüglich Versandkosten. Vorzugspreis für Mitglieder der Deutschen Gesellschaft für Medizinische Dokumentation und Statistik sowie der Deutschen Region der Internationalen Biometrischen Gesellschaft DM 194,- zuzüglich Versandkosten. Bestellungen nehmen jede Buchhandlung sowie die Verlage entgegen.

# BIOMETRIE UND INFORMATIK

IN MEDIZIN UND BIOLOGIE

---

## Bd. 20 (1989) Inhaltsverzeichnis

Vorwort	1
Editorial	2, 129
BAUER, P.: Multistage testing with adaptive designs	130
HAUX, R.: Statistische Expertensysteme	3
HILGERS, R. A. and NEUMANN, N.: An exact nonparametric randomization test for censored data	66
HOFMANN, E. und PRECHT, M.: Besondere Anlagetypen und ihre varianz-analytische Auswertung	74
PRÖSELER, M. und WALOSZCZYK, K.: Ein Modell zur rechnergestützten Sortenwahl	101
SCHUMACHER, M.: Frailty-Modelle und ihre Anwendung in der Medizin	89
TAFEIT, E., LABACK, O., PALETTA, B. und POGGLITSCH, H.: Ein Algorithmus zur individuellen Diätplanerstellung am Personal-Computer	116
THÖNI, H.: Fehlstellen in orthogonal geplanten Versuchen: geschlossene Lösung zur Berechnung von Ersatzwerten und zur exakten Berechnung der Standardfehler der Prüfglied-Differenzen	149
WELLEK, S.: Computing exact p-values in Page's nonparametric test against trend	163
Fallbeispiele	
ULM, K., SCHMOOR, C., SAUERBREI, W., KEMMLER, G., AYDEMIR, Ü. und MÜLLER, B.: Strategien zur Auswertung einer Therapiestudie mit Überlebenszeit als Zielkriterium	171
Buchbesprechungen	86, 128, 206
Persönliches	87
Danksagung	208



## Vorwort

Nach fast 20 Jahren war es geboten, den Inhalt und auch das Erscheinungsbild der Zeitschrift „EDV in Medizin und Biologie“ einmal kritisch zu überdenken. Intensive Gespräche in verschiedenen Bereichen und in verschiedenen Gremien haben dazu geführt, die Zeitschrift „EDV in Medizin und Biologie“ nun ab 1989 in neuer Form als

### **BIOMETRIE und INFORMATIK in Medizin und Biologie**

herauszugeben.

Ausschlaggebend auch für die Titelländerung waren u. a. zwei Gründe:

- der Anteil an statistisch- bzw. biometrisch-orientierten Arbeiten nahm ständig zu und
- die reinen Computeranwendungen traten immer mehr zurück gegenüber Arbeiten aus dem Bereich der Informatik.

Mit der Titelländerung haben wir eine Reihe inhaltlicher und formaler Änderungen verbunden, von denen wir hoffen, daß sie langfristig die Akzeptanz und Verbreitung der Zeitschrift fördern werden. In diesem Zusammenhang wurde auch der Herausgeber- und Mitarbeiterstab vergrößert, um den zu erwartenden Anforderungen Rechnung zu tragen.

„BIOMETRIE und INFORMATIK in Medizin und Biologie“ ist eine deutsch/englischsprachige Zeitschrift, die theoretische und anwendungsbezogene Arbeiten aus dem Gesamtgebiet der Biometrie und Informatik in Medizin, Biologie, Agrarwissenschaften und anderen Anwendungsgebieten enthält. Die Zeitschrift veröffentlicht sowohl Beiträge über neu entwickelte Ansätze als auch über neue Anwendungsmöglichkeiten bekannter Methoden. Zusätzlich soll dem tutoriellen Aspekt durch Übersichtsartikel Rechnung getragen werden. Der Bezug zu einem der Anwendungsgebiete muß dabei jedoch hergestellt werden.

Die Qualität der Beiträge wird durch ein sorgfältiges Begutachtungsverfahren sichergestellt.

Die Herausgeber und die Schriftleitung danken den beiden Verlagen für das Verständnis bei der Umstellung der Zeitschrift.

Die Herausgeber und die Schriftleitung

Frühjahr 1989

## Editorial

In diesem ersten Heft der „BIOMETRIE und INFORMATIK in Medizin und Biologie“ wurde als Leitthema

### „Statistische Expertensysteme“

gewählt. Herr R. HAUX konnte gewonnen werden, ein Referat, das er 1987 auf dem Biometrischen Seminar der Region Österreich-Schweiz der Internationalen Biometrischen Gesellschaft gehalten hat, in überarbeiteter und ergänzter schriftlicher Fassung zur Verfügung zu stellen.

Das Thema erscheint uns für das erste in neuer Form erscheinende Heft besonders geeignet, weil es wie kaum ein anderes beide – manchmal so disjunkt erscheinende – Hauptgebiete dieser Zeitschrift verknüpft. Außerdem ist es derzeit aktuell: Erste Entwicklungen liegen vor oder zeichnen sich ab, und die Tragfähigkeit und die Entwicklungsmöglichkeiten in diesem neuen Bereich werden zur Zeit ausführlich diskutiert. Auf der einen Seite herrscht Optimismus, auf der anderen Seite steht deutliche Ablehnung, eine Situation, die sicherlich zum Teil auch darin begründet liegt, daß unterschiedliche Definitionen und Vorstellungen von der Konzeption und der Leistungsfähigkeit statistischer Expertensysteme bestehen. Somit erscheint eine Einführung in Grundbegriffe, existierende Beispiele, Stand der Entwicklung und eine Diskussion des möglichen Nutzens solcher Systeme von großem Interesse.

Die Präsentation des Hauptthemas in Form einer einführenden Übersicht und einer Folge von Kommentaren soll exemplarisch verstanden werden: Es wird angestrebt, auch weiterhin Einführungs- und Übersichtsartikel zu veröffentlichen, die aber keineswegs immer einen abgerundeten und endgültigen Charakter zu haben brauchen, sondern das Lückenhafte und Unfertige kann auch durch zusätzliche Kommentare anderer Autoren aufgezeigt und ergänzt werden.

## Statistische Expertensysteme\*

R. Haux

### Zusammenfassung

*Die Arbeit enthält einen Überblick über den derzeitigen Stand der Entwicklung Statistischer Expertensysteme. Es werden die für die Erörterung notwendigen Grundbegriffe zu Expertensystemen eingeführt. Nach einer kurzen Beschreibung existierender Systeme wird Stellung zum Stand der Entwicklung bezogen und der Nutzen Statistischer Expertensysteme für die Statistik und für die Biometrie diskutiert.*

### Schlüsselwörter

*Statistische Expertensysteme; Computational Statistics*

### Summary

*The intention of the paper is to describe the state of the art concerning the development of statistical expert systems. First the basic notation and basic definitions needed for the discussion of such expert systems are introduced. After a brief description of existing systems, the author comments on the development of statistical expert systems and points out some methodological problems. Finally the usefulness of statistical expert systems for statistics and for biometry is discussed.*

### Key words

*statistical expert systems; computational statistics*

### 1 Einleitung

STREITBERG schreibt in einer bewußt provokativ gehaltenen Arbeit über Expertensysteme in der Statistik ([113], [114]):

- »(1) Es gibt keine statistischen Expertensysteme
- (2) Es wird auch niemals solche Systeme geben
- (3) Da es sie nicht geben kann
- (4) Dennoch ist es sinnvoll, an den Problemstellungen derartiger Systeme zu arbeiten«

\* Herrn Professor Dr. med. HERBERT IMMICH zum siebzigsten Geburtstag gewidmet. (Erweiterte und aktualisierte Fassung eines Vortrages, gehalten auf dem Biometrischen Seminar der Region Österreich-Schweiz (ROeS) der Internationalen Biometrischen Gesellschaft, 21.–25. 9. 1987, Locarno, Schweiz.)

([113], S. 90) und versucht diese Aussagen vor allem durch Komplexitätstheoretische Argumente zu belegen.

Die vorliegende Arbeit enthält eine Übersicht über den Stand der Entwicklung Statistischer Expertensysteme und eine kritische Wertung dieses Standes. Die Arbeit enthält auch eine Stellungnahme zu den Thesen STREITBERGS.

Zum Aufbau der Arbeit: In Abschnitt 2 werden einige Grundbegriffe zu Expertensystemen eingeführt, soweit sie für diese Arbeit wichtig sind. Dieser Abschnitt ist zum Teil angelehnt an [64], Kapitel 2 und an [65]. Er ist so ausgearbeitet, daß er auch für Biometriker ohne weitergehende Kenntnisse auf dem Fachgebiet der Expertensysteme lesbar sein sollte. Abschnitt 3 vermittelt einen ersten Überblick über den Stand der Entwicklung von Expertensystemen in der Statistik. Der Abschnitt enthält außerdem Vorschläge für Begriffsbestimmungen im Zusammenhang mit Statistischen Expertensystemen. Er wird ergänzt durch die Beschreibung einiger Statistischer Expertensysteme in Abschnitt 4. Da die Entwicklung von Expertensystemen in der Statistik praktisch nicht mehr zu überblicken ist, kann und soll kein Anspruch auf Vollständigkeit bei der in den Abschnitten 3 und 4 zitierten Literatur erhoben werden. In Abschnitt 5 nimmt der Verfasser zu einzelnen Problemen bei dem Entwurf und der Anwendung Statistischer Expertensysteme – und soweit hier notwendig, bei Expertensystemen allgemein – Stellung. Es folgt eine abschließende Diskussion mit einer Stellungnahme zu den Thesen STREITBERGS.

Während der Verfasser in den Abschnitten 2 bis 4 den Stand der Entwicklung Statistischer Expertensysteme möglichst wertfrei zu beschreiben versucht, wertet er in Abschnitt 5 Aspekte dieser Entwicklung und der Anwendung solcher Systeme. Diese notwendigerweise subjektive und teilweise kritische Wertung mag in dem einen oder anderen Punkt zu weit gehen. Sollte dies der Fall sein, so hofft der Verfasser mit dieser Kritik zumindest zu einer lebhaften Diskussion über die Entwicklung und Anwendung von Statistischen Expertensystemen beigetragen zu haben – zum Nutzen der Statistik und der Biometrie. Manche Ausführungen zu Expertensystemen und zu Statistischen Expertensystemen müssen aus Gründen der besseren Lesbarkeit verkürzt und auch vereinfacht dargestellt werden – allerdings unter Angabe weiterführender Literatur. Auch hierfür bittet der Verfasser um Nachsicht. Zum Verfasser sei noch gesagt, daß er zwar auch praktische Erfahrungen bei der Entwicklung von Expertensystemen zur Diagnose- und Therapieunterstützung und bei der statistischen Planung, Durchführung und Auswertung von Studien sammeln konnte, daß er jedoch keine praktische Erfahrung bei der Entwicklung eines Statistischen Expertensystems hat. Die Sicht des Verfassers ist außerdem geprägt durch die Anwendung statistischer Methoden in der Medizin.

## 2 Expertensysteme

### 2.1 Begriffsbestimmung

Wir wollen eine Maschine (zum Beispiel eine TURING-Maschine, einen deterministischen endlichen Automaten) genau dann als *Expertensystem* bezeichnen, wenn sie Funktionen, die für einen Experten spezifisch sind, übernehmen kann ([64], S. 15). Da man solche Maschinen üblicherweise auf »universellen« (zum Beispiel von NEUMANN) Rechnern durch »spezielle« Programme realisiert, läßt sich auch in der obigen Begriffsbestimmung der Begriff Maschine durch Programm ersetzen, ohne allzu unpräzise zu werden. Unter (spezifischen) Funktionen beziehungsweise Tätigkeiten von Experten verstehen wir diejenigen Tätigkeiten, die, außer von Expertensystemen, nur von Experten übernommen werden können. Ein Beispiel für eine spezifische Expertentätigkeit ist etwa die Diagnose- und Therapiefindung in der Medizin.

In der Literatur werden Expertensysteme auf unterschiedliche Weise definiert. In einschlägigen Fachlexika bezeichnet man sie beispielsweise als »Computersystem, das auf

einem speziellen Wissensgebiet die Kompetenz von menschlichen Experten besitzt und als Beratungs- und Problemlösungssystem eingesetzt wird« ([107], S. 219) oder als »Programm-System, das ›Wissen‹ über ein spezielles Gebiet speichert und ansammelt, aus Wissen Schlußfolgerungen zieht und zu konkreten Problemen des Gebietes Lösungen anbietet« ([83], S. 222). In der Fachliteratur werden Expertensysteme eingeführt als »Programme, mit denen das Spezialwissen und die Schlußfolgerungsfähigkeit qualifizierter Fachleute auf eng begrenzten Aufgabengebieten nachgebildet werden soll« ([100], S. 2), als »Computerprogramme, die Fähigkeiten von Experten simulieren sollen« ([99], S. 8) oder als (im Singular) »intelligentes Computerprogramm, das Wissen und Inferenzverfahren benutzt, um Probleme zu lösen, die immerhin so schwierig sind, daß ihre Lösung ein beträchtliches menschliches Fachwissen erfordert« ([57], S. 3). In manchen Lehrbüchern wird ein Experte als jemand bezeichnet, »der Fakten und Regeln« . . . »aus einem bestimmten Wissensgebiet besser beherrscht und besser zu manipulieren versteht als die Mehrzahl anderer Menschen. Ein Expertensystem ist demnach ein Computersystem, welches gebietsspezifisches Expertenwissen speichern, verwalten, gezielt auswerten und zu Auskünften an einen Benutzer oder zur Abwicklung bestimmter Aufgaben« . . . »nutzen kann« ([109], S. 1, vgl. auch [108], S. 5). Es gibt allerdings auch namhafte Vertreter auf dem Fachgebiet Expertensysteme, die vorschlagen, auf eine Definition des Begriffs Expertensystem ganz zu verzichten.

Das Fachgebiet Expertensysteme wird außerdem häufig dem Fachgebiet *Künstliche Intelligenz* (KI beziehungsweise englisch AI abgekürzt; [22], [94], [122]) zu- und untergeordnet, einem Fachgebiet, welches »bisher dem Menschen vorbehaltenen Verhaltensweisen«, die Intelligenz erfordern, untersucht, »indem sie sie auf dem Rechner simuliert und naturwissenschaftlicher Betrachtungsweise und damit ingenieurmäßiger Verwendung zugänglich macht« ([5], S. 1). Beispiele für solche Verhaltensweisen sind das Planen von Reisen, das Verstehen von gesprochener Sprache, das Beweisen mathematischer Sätze, das Stellen einer medizinischen Diagnose und das Sehen und Erkennen von Gegenständen ([5], S. 1). Hauptdisziplinen der Künstlichen Intelligenz sind nach [5] natürlichsprachliche Systeme, Deduktionssysteme, Robotik, Bildverstehen, Cognitive Science und Expertensysteme.

## 2.2 Funktionsweise

Expertensysteme haben bestimmte *Anwendungsgebiete*, in denen sie eingesetzt werden können, zum Beispiel die oben erwähnte Diagnose- und Therapiefindung in der Medizin. Es lassen sich außerdem bei Expertensystemen üblicherweise zwei *Personengruppen* unterscheiden, die in unterschiedlicher Weise mit einem Expertensystem arbeiten: die *Benutzer* eines Expertensystems und die *Wissensverwalter*.

Die Funktionsweise aus der Sicht eines Benutzers ist grob folgende: Es wird ein konkretes Problem beschrieben (›Eingabe‹), für das unter Verwendung von Expertenwissen eine Problemlösung (›Ausgabe‹) vorgeschlagen wird. Beispielsweise beschreibt ein Benutzer den Zustand eines Patienten anhand bestimmter Befunde und möchte einen Diagnosevorschlag erhalten.

Die Funktionsweise aus der Sicht eines Wissensverwalters dagegen ist: Es wird (Experten-) Wissen über ein bestimmtes Anwendungsgebiet beziehungsweise über Strategien zur Problemlösung für das Anwendungsgebiet spezifiziert (›Eingabe‹) beziehungsweise die Spezifikation gelesen oder überprüft (›Ausgabe‹). Beispielsweise spezifiziert ein Wissensverwalter, bei welchen Befunden eine bestimmte Diagnose vorgeschlagen werden soll.

Die Vorgehensweise, wie man von einem Problem zu einem Problemlösungsvorschlag gelangt, bezeichnet man als *Inferenzmechanismus*. Einfache Inferenzmechanismen basieren beispielsweise unmittelbar auf (deterministischen) Entscheidungstabellen (z. B. [75]), auf bedingten Wahrscheinlichkeiten (Satz von BAYES, z. B. [1]) oder auf diskriminanzanalytischen Verfahren (z. B. [116]). Eine Übersicht über verschiedene Arten von Inferenzmecha-

nismen befindet sich in [110]. Weitere Inferenzmechanismen werden in Abschnitt 2.5 beschrieben.

Neben Benutzern und Wissensverwaltern läßt sich im Zusammenhang mit Expertensystemen noch eine dritte Personengruppe nennen: die (Expertensystem-) *Konstrukteure*. Diese drei Personengruppen müssen nicht notwendigerweise disjunkt sein.

### 2.3 Qualifikation von Benutzern und Wissensverwaltern

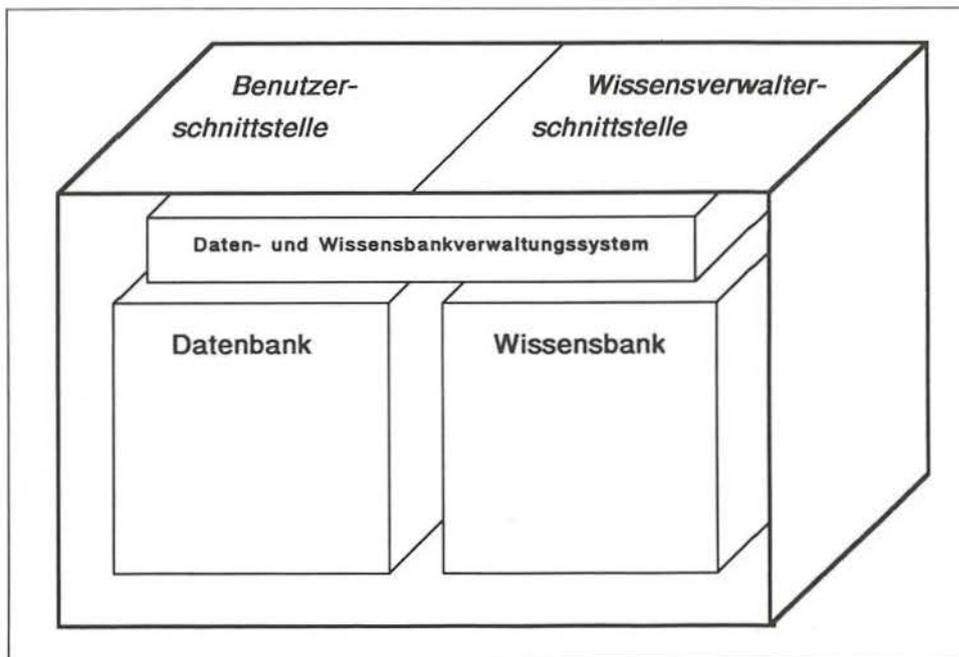
Bezüglich der Qualifikation von Benutzern und Wissensverwaltern lassen sich unterscheiden:

- der *Laie*, der, bezogen auf das Anwendungsgebiet, keine Fachkenntnisse hat, und
- der *Experte*, der, bezogen auf das Anwendungsgebiet, umfangreiches (Erfahrungs-)Wissen besitzt.

Laie einerseits und Experte andererseits bilden zwei Pole, zwischen denen es eine Vielzahl von Spezialisierungsgraden gibt. So könnte man etwa eine Person, die Grundkenntnisse auf einem Anwendungsgebiet besitzt und in der Lage ist, Probleme bestimmten Problembereichen des Anwendungsgebiets richtig zuzuordnen, als *Fachperson* bezeichnen.

Bei Wissensverwaltern sollte es sich grundsätzlich um Experten des Anwendungsgebietes handeln. Bei Benutzern sind alle Qualifikationsstufen möglich. Expertensysteme, die bezogen auf das Anwendungsgebiet, Laien als Benutzer haben, sollen Experten ersetzen. Expertensysteme, die, bezogen auf das Anwendungsgebiet, Experten als Benutzer haben, dienen der Unterstützung von Experten. Das Wissen von Benutzern kann also – muß jedoch nicht – geringer als das der Wissensverwalter sein, natürlich immer nur bezogen auf das jeweilige Anwendungsgebiet eines Expertensystems! Systeme, bei denen das ›Wissensgefälle‹ (insbesondere im Hinblick auf Erfahrungswissen) zwischen Benutzern und Wissens-

Abbildung 1. Aufbau eines Expertensystems.



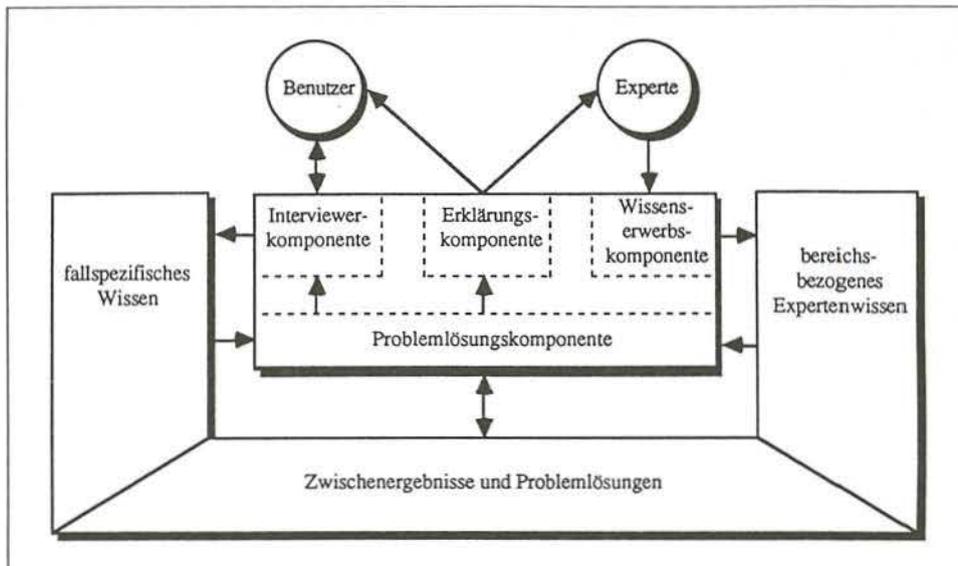


Abbildung 2. Aufbau eines Expertensystems nach [100], S. 13.

verwalten nicht allzu groß ist, dienen dem Wissenstransfer. Man versucht sozusagen, das vorhandene Wissensgefälle auszugleichen. Beispielsweise können Expertensysteme zur Diagnose- und Therapieunterstützung für auf dem Anwendungsgebiet noch unerfahrene Assistenzärzte als Benutzer sein und auf das Anwendungsgebiete spezialisierte, erfahrene Oberärzte oder Chefärzte als Wissensverwalter haben.

#### 2.4 Architektur

Der Aufbau eines Expertensystems läßt sich folgendermaßen gliedern: Ein Expertensystem besteht aus einer Wissensbank und aus einer Datenbank. Die Datenbank und die Wissensbank werden aufgebaut und verwaltet durch ein Daten- und Wissensbankverwaltungssystem. Jede dieser drei Komponenten kann mehr oder weniger stark ausgeprägt sein. Eine schematische Darstellung befindet sich in Abbildung 1.

In der Datenbank sind beispielsweise die Ausprägungen von Befunden eines Patienten enthalten, in der Wissensbank Wissen über Krankheiten.

Das Daten- und Wissensbankverwaltungssystem benutzt, nach bestimmten Vorgaben, das Wissen aus der Wissensbank zur Problemlösung für Daten aus der Datenbank. Wichtige Ausgaben sind, neben der Datenbankverwaltung und der Wissensbankverwaltung, das Ableiten und Vorschlagen einer Problemlösung, die Erklärung der Problemlösung sowie die Ein-/Ausgabe von Daten an die Benutzer über die Benutzerschnittstelle und von Wissen an die Wissensverwalter über die Wissensverwalterschnittstelle. Die Architekturkomponente des Daten- und Wissensbankverwaltungssystems zum Vorschlagen einer Problemlösung bezeichnet man auch als Inferenzkomponente, die zur Erklärung eines Problemlösungsvorschlags als Erklärungskomponente. Beide Komponenten hängen eng miteinander zusammen.

In der Literatur wird der allgemeine Aufbau eines Expertensystems häufig in der in Abbildung 2 dargestellten Weise beschrieben. Das dort enthaltene fallspezifische Wissen

entspricht den Daten des Benutzers, das bereichsbezogene Wissen dem Wissen des Wissensverwalters, der Experte dem Wissensverwalter.

Das Daten- und Wissensbankverwaltungssystem, sozusagen ein Expertensystem mit ›leerer‹ Datenbank und ›leerer‹ Wissensbank, wird häufig als *Expertensystem-Shell* (›expert system shell‹) bezeichnet.

Je nach Anwendungsgebiet können die Architekturkomponenten eines Expertensystems weiter verfeinert werden.

### 2.5 Repräsentationsformen

Die Daten zu einem konkreten Problem und das Wissen über ein bestimmtes Anwendungsgebiet können in einem Expertensystem nicht unmittelbar, an sich, verwendet werden. Man muß – da ein Expertensystem ja eine Maschine ist – Daten und Wissen formal repräsentieren. Formale Repräsentationsarten bezeichnet man im Zusammenhang mit Daten häufig als *Datenmodelle*, im Zusammenhang mit Wissen als *Wissensrepräsentationsformen*. Letztendlich muß man sowohl formal repräsentierte Daten als auch formal repräsentiertes Wissen als Datenstrukturen darstellen. Die Spezifikation von Daten und Wissen mittels bestimmter Datenmodelle und Wissensrepräsentationsformen wird über dafür möglichst geeignet gewählte *formale Sprachen* vorgenommen (unmittelbar über die jeweilige Benutzer- oder Wissensverwalterschnittstelle beziehungsweise mittelbar über allgemein verwendbare Sprachen wie die funktionale Sprache LISP, die Sprache PROLOG, die objektorientierte Sprache SMALLTALK, aber auch über imperative Sprachen wie PASCAL). Die wohl am meisten bekannte und zitierte Wissensrepräsentationsform ist die Repräsentation über sogenannte *Produktionsregeln*. Mit ihnen repräsentiert man Wissen als endliche Menge von ›wenn – dann‹ Konstrukten, etwa in der Art

WENN ›Bedingung wahr ist‹ (Testkomponente)  
DANN ›gilt Schlußfolgerung‹ (Aktionskomponente).

Beispielsweise lassen sich (Produktions-)Regeln wie

WENN	Hepatitis-B-Antigen-Titer hoch	DANN	Hepatitis B,
WENN	Hepatitis-A-IgM-Antikörper nachgewiesen	DANN	Hepatitis A,
WENN	Hepatitis B und Hepatitis A	DANN	Hepatitis-Doppelinfection

spezifizieren. Durch welche Meßverfahren Antigene und Antikörper nachgewiesen werden können und was in diesem Zusammenhang ›hoch‹ bedeutet, sei hier vernachlässigt. Ein ausführlicheres Beispiel einer formal spezifizierten Wissensbank zur Hepatitis-Diagnostik, welches mit dem Daten- und Wissensbankverwaltungssystem EXPERT ([120]) realisiert wurde, befindet sich in [65], Teil 2.

Bei solchen regelbasierten Expertensystemen wird üblicherweise durch Eingabe von *Fakten* durch den Benutzer (Problembeschreibung, zum Beispiel daß der Hepatitis-B Antigen-Titer eines Patienten hoch ist) von dem Inferenzmechanismus anhand der in der Wissensbank enthaltenen *Regeln* ein Problemlösungsvorschlag abgeleitet (zum Beispiel daß der Patient eine Hepatitis-B-Infektion hat).

Die formale Repräsentation über Produktionsregeln läßt sich weiter verfeinern, zum Beispiel indem man Gewichte für die beiden Komponenten, sogenannte *Sicherheitsfaktoren*, vergibt.

Weitere Wissensrepräsentationsformen sind, u. a., sogenannte Frames (›Rahmen‹ für die strukturierte Beschreibung von Wissensobjekten) beziehungsweise allgemein aussagenlogische oder prädikatenlogische und darauf basierende modallogische Ausdrücke (Näheres vgl. [64] S. 31, 70–100, [70], [94], [100], S. 15–42, [103], [119]).

Es dürfte unmittelbar klar sein, daß vor allem die jeweilige Wissensrepräsentationsform eng mit dem jeweiligen Inferenzmechanismus zusammenhängt. Im Zusammenhang mit (produktions-)regelbasierten Expertensystemen unterscheidet man hier zwischen zielorien-

tierten Inferenzmechanismen (ausgehend von einem angeforderten Problemlösungsvorschlag versucht man das zu untersuchende Problem in Teilprobleme zu zerlegen und dann durch Erfragen geeigneter Daten die vermutete Problemlösung zu bestätigen) und datenorientierten Inferenzmechanismen (ausgehend von eingegebenen Daten versucht man eine Problemlösung vorzuschlagen oder weitere Daten zu erfragen). Beide Inferenzmechanismen stellen ebenfalls nur zwei, wenn auch typische Arten von Inferenzmechanismen dar.

## 2.6 Literatur

Eine zwar kurze, aber angenehm unkonventionelle Einführung zu Expertensystemen befindet sich – vor allem für Informatiker – in [103]. Umfassende Einführungen befinden sich u. a. in [22], [57], [70], [94], [99], [100], [108], [109], [119], [120] und [122]. Ein sorgfältig editiertes Buch mit wichtigen Arbeiten auf dem Gebiet der Expertensysteme in der Medizin ist [24]. Für regelbasierte Expertensysteme sei auf [12] verwiesen.

Wie bereits erwähnt, orientiert sich der vorliegende Abschnitt 2 an [64], Kapitel 2 und an [65]. Dort befinden sich auch nähere Erläuterungen zum Inhalt dieser sehr kurz gehaltenen Einführung, weitere Literaturangaben sowie Anwendungsbeispiele und Verweise auf existierende Systeme.

## 3 Expertensysteme in der Statistik

### 3.1 Begriffsbestimmung

Ein Expertensystem wollen wir genau dann als *Statistisches Expertensystem* bezeichnen, wenn es spezifische Expertentätigkeiten in der Statistik oder in einem Teilgebiet hiervon übernehmen kann.

### 3.2 Anwendungsgebiete

Das Anwendungsgebiet Statistik und Teilgebiete hiervon lassen sich charakterisieren durch:

- die *Art der statistischen Auswertung*: zum Beispiel konfirmatorisch, explorativ;
- die *Phase der statistischen Auswertung*: zum Beispiel Studienplanung und Auswahl statistischer Methoden (zum Beispiel Test-, Schätzverfahren), Studiauswertung und Ergebnisinterpretation;
- das *statistische Methodenspektrum*: zum Beispiel Kontingenztafelanalyse, lineare Rangtests;
- das *statistische Anwendungsgebiet*: zum Beispiel Ökologie, Medizin, Chemie, Landwirtschaft.

Zu den Aufgaben, die von Statistischen Expertensystemen übernommen werden könnten, zählen zum Beispiel die (Unterstützung bei der) Auswahl statistischer Testverfahren für ein bestimmtes statistisches Anwendungsgebiet oder die (Unterstützung bei der) explorativen Datenanalyse mit regressionsanalytischen Methoden.

Eine Diskussion darüber, welche Aufgaben in welchen Teilgebieten der Statistik und für welche Benutzer durch Statistische Expertensysteme übernommen werden können, befindet sich in [62], S. 179–187. In Abhängigkeit von der Funktionsweise beziehungsweise von den jeweiligen Aufgaben von Statistischen Expertensystemen wird vorgeschlagen, solche Systeme auch als wissensbasiertes Beratungssystem (knowledge-based consultation system; [42]), wissensbasierte Schnittstelle (knowledge-based front end; [127], [91]), Statistisches Wissensvergrößerungssystem (statistical knowledge enhancement system; [56]), Expertenmodellierungssystem (expert-modeling system; [20]) und Expertenleistungssystem (expert-performance system; [20]) zu bezeichnen.

Statistik als Anwendungsgebiet von Expertensystemen wird hier also nicht nur als die

Mathematische Statistik verstanden, sondern auch und vor allem als Angewandte Statistik, bei der man versucht, Probleme eines statistischen Anwendungsgebietes zu modellieren, statistische Methoden hierfür zu verwenden und Hilfe bei der Durchführung von Studien und bei der Interpretation der gewonnenen Ergebnisse zu leisten.

### 3.3 Personengruppen

Wissensverwalter sind Statistiker, die, bezogen auf das Anwendungsgebiet, umfangreiches (Erfahrungs-)Wissen besitzen sollten, die also Experten auf diesem Anwendungsgebiet sein sollten.

Bei den Benutzern können wir bei Statistischen Expertensystemen vor allem drei Gruppen unterscheiden:

- den *statistischen Laien*, zum Beispiel einen Arzt, der sich noch nicht mit statistischen Methoden befaßt hat;
- die *statistische Fachperson*, die in dem jeweiligen Anwendungsgebiet Grundkenntnisse besitzt;
- den auf dem Anwendungsgebiet versierten *statistischen Experten*.

Laien auf der einen Seite und Experte auf der anderen Seite stellen wiederum zwei Pole dar.

### 3.4 Statistische Expertensysteme und Statistische Auswertungssysteme

Statistische Expertensysteme sind, bezogen auf deren Grundaufbau, verwandt mit *Statistischen Auswertungssystemen* (wie BMDP, DATA DESK, GLIM, S, SAS, SPSS; vgl. hierzu [3], [15], [35], [59], [117]), die sich über die drei Architekturkomponenten *Datenbank*, *Programmbank* (bzw. *Methodenbank*) und *Daten- und Programmbankverwaltungssystem* beschreiben lassen. Bei Statistischen Expertensystemen versucht man auch, Wissen über die in Datenbank und Programmbank gespeicherten Daten und Programme formal zu repräsentieren und bei der Problemlösung zu verwenden. Die eigentlichen Daten und Programme werden dann möglicherweise über Schnittstellen angesprochen (z. B. [27]).

Eine verfeinerte, mehrschichtige Architektursicht für Statistische Expertensysteme, die auf das Anwendungsgebiet Statistik zugeschnitten ist, befindet sich in [32], ein Vergleich verschiedener Arten von Benutzerschnittstellen in [126].

### 3.5 Literatur

Die folgenden Bände – [39], [62] und [97] – befassen sich schwerpunktmäßig mit Statistischen Expertensystemen: Der erste basiert auf einer Tagung, die im April 1985 in Princeton (N.J.) stattfand [39], der zweite auf einer Tagung, die im November 1985 in Aachen ausgerichtet wurde [62]. In diesen drei Bänden dürfte der derzeitige Stand der Entwicklung Statistischer Expertensysteme wohl am umfassendsten enthalten sein. Verwiesen sei hier exemplarisch auf [40], [41], [54] in [39], auf [76], [115], [125], [127] in [62] und auf [90] in [97]. Als Nachfolger der ›Princeton-Konferenz‹ fand im Januar 1989 der ›2nd International Workshop on Artificial Intelligence and Statistics‹ in Fort Lauderdale, Florida, statt. Ein Tagungsband dürfte hier ebenfalls zu erwarten sein.

Weitere Angaben befinden sich in den COMPSTAT-Tagungsbänden, zum Beispiel [50] in [17], [45], [98] in [69], [37], [96] in [27], [86], [91], [121] in [31], in den ›Interface‹-Bänden, zum Beispiel [18] in [30], [44] in [72], [93], [124] in [9]. Auf den SoftStat Konferenzen über die wissenschaftliche Anwendung von Statistik-Software ([33], [34]) gab es ebenfalls mehrere Sitzungen über Statistische Expertensysteme, verwiesen sei hier exemplarisch auf [92], [126]. Weitere Literatur befindet sich u. a. im Statistical Software Newsletter (z. B. [32], [101], [114]), den Bänden des ISI (International Statistical Institute) und der IASC (International Association for Statistical Computing), zum Beispiel [19], [21], [42] sowie in den Bänden weiterer statistischer Gesellschaften (z. B. [95]). In statistischen Zeitschriften und

Büchern wird dieses Thema ausführlicher behandelt (z. B. [29], [36], [43], [49], [51], [52], [56], [81]) als in Zeitschriften und Büchern, die sich überwiegend mit Themen aus der Informatik befassen (z. B. [38], [48], [53], [63]). Literatur zu Statistischen Expertensystemen wird an sehr unterschiedlichen Stellen veröffentlicht. Sie läßt sich mittlerweile nur noch schwer, wenn überhaupt, überblicken. Eine Liste von Wissenschaftlern, die sich mit der Entwicklung Statistischer Expertensysteme befassen, wurde von D. J. HAND zusammengestellt ([55]).

## 4 Stand der Entwicklung bei Statistischen Expertensystemen

### 4.1 Vorbemerkungen

Überlegungen, in der Mathematischen oder Angewandten Statistik durch den Einsatz von Rechnern nicht nur bei dem bloßen »numerischen« Ausrechnen von Teststatistiken und ähnlichem Unterstützung zu erhalten, sind älter, als es der Begriff Statistisches Expertensystem vermuten läßt. Erinnerung sei an dieser Stelle vor allem an eine Arbeit von NELDER [88], aber auch, exemplarisch, an [10], [58], [77], [89].

NELDER schreibt in der Arbeit »intelligent programs, the next stage in statistical computing« ([88], S. 79): »Current computer programs for statistical computing are largely unintelligent in that they do not use the data to check if the assumptions underlying the statistical procedure being used are approximately satisfied« . . . »Most current programs will not give the user any explicit warning that anything is wrong, so the uncritical user will not know if the results are wrong or not.« . . . »The amount of uncritical use of standard procedures is enormous and, in my view, is tending to bring the subject of statistics into disrepute«. Es geht NELDER darum, die Gefahr einer falschen Anwendung von Programmen für die statistische Datenanalyse zu vermindern.

Vier Jahre nach Erscheinen von NELDERs Arbeit verwendet CHAMBERS ([18]) zum ersten Mal (nach Wissen des Verfassers) den Begriff Expertensystem in Verbindung mit dem Anwendungsgebiet Statistik (der Begriff wurde – nebenbei bemerkt – in der Medizin bereits ein knappes Jahrzehnt früher benutzt). Wie in zahlreichen anderen Arbeiten über Statistische Expertensysteme auch rechtfertigt CHAMBERS seine Arbeiten auf diesem Gebiet ebenfalls mit der Gefahr der mißbräuchlichen Anwendung statistischer Software.

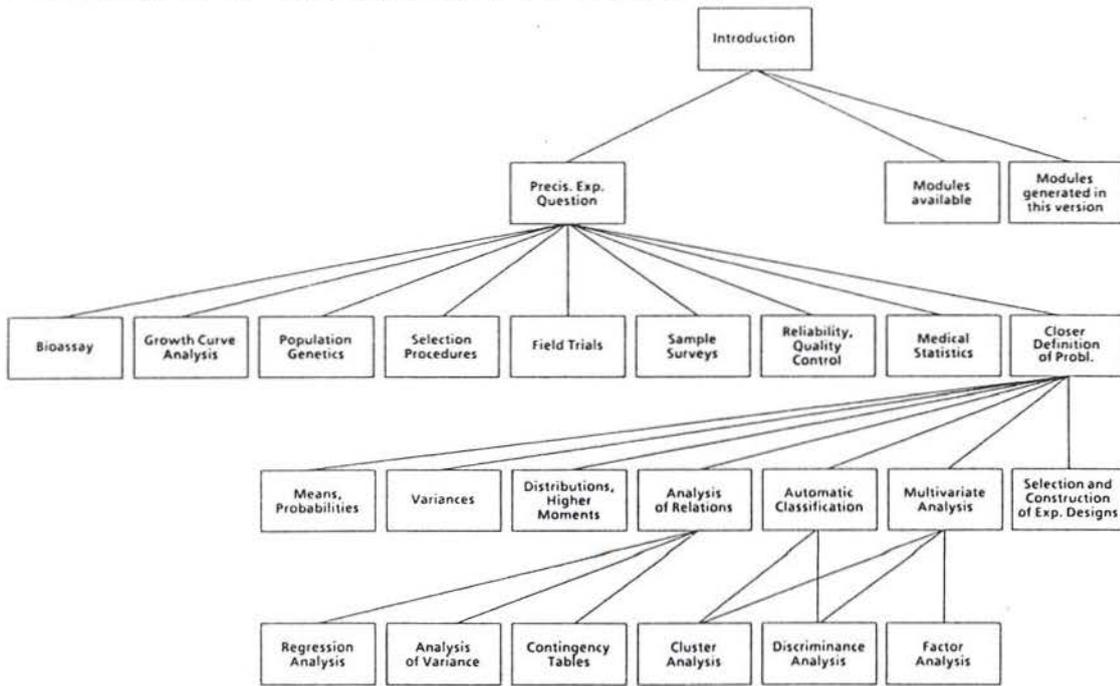
CHAMBERS schreibt ([19], S. 101): »We need diagnostic techniques over a broad range of analytical methods, so that these methods may be used by non-professionals with some chance of alerting the client to possible dangers.« . . . »Because data analysis will take place far from expert statistical help, the challenges must be faced if quality analysis is possible.« (vgl. auch zum Beispiel [49]). CHAMBERS argumentiert also, daß solche Systeme einen (nicht vorhandenen) statistischen Experten ersetzen sollen.

Statistische Expertensysteme lassen sich als Weiterentwicklung der Statistischen Auswertungssysteme sehen. Neben einer für den Benutzer möglichst einfach aufzurufenden und numerisch richtig gerechneten statistischen Datenanalyse sollen solche Systeme in verstärktem Maße darauf achten, daß das verwendete statistische Verfahren oder das verwendete statistische Modell für das zu untersuchende Problem angemessen ist.

### 4.2 Existierende Systeme

Im Gegensatz zu der Vielfalt an Veröffentlichungen über Statistische Expertensysteme ist die Zahl existierender Expertensysteme, welche Wissen aus dem Gebiet der Statistik verwenden, gering. Drei Statistische Expertensysteme, GLIMPSE, PANOS und REX, sollen im folgenden kurz beschrieben werden. Allerdings muß hinzugefügt werden, daß in der letzten Zeit häufiger über realisierte Systeme berichtet wird. Erwähnt seien an dieser Stelle die Systeme CADEMO ([101], [102]), dessen schematische Gliederung der Wissens-

Abbildung 3. Schematische Gliederung der Wissensbank von CADEMO (aus [101], S. 108).



bank in Abbildung 3 enthalten ist, KENS [56] sowie [4], [16], [26], [27], [29], [37], [50], [81] und [115].

Viele dieser Systeme sind prototypische Systeme oder Systeme, die zwar lauffähig sind, die jedoch nur in ersten, einfachen Grundausstufen existieren. Außer CADEMO wird keines dieser Systeme kommerziell vertrieben, allerdings ist dies auch für GLIMPSE angekündigt. Manche dieser Statistischen Expertensysteme basieren auf Expertensystem-Shells, zum Beispiel EXPRESS ([16]) auf EXPERT, viele verfügen wie bereits erwähnt über Schnittstellen zu Statistischen Auswertungssystemen.

#### 4.3 GLIMPSE

GLIMPSE steht für *GLIM + Prolog + Statistical Expertise*. Es wurde unter maßgeblicher Beteiligung von J. A. NELDER am Imperial College of Science and Technology, London, als komfortable Benutzerschnittstelle für das Statistische Auswertungssystem GLIM (*Generalized Linear Interactive Modelling*) entwickelt ([90], [91], [92], [93], [127]). Der Benutzer – eine statistische Fachperson, die sich zumindest etwas mit generalisierten linearen Modellen auskennt – soll unterstützt werden bei der (explorativen) Zuordnung eines zu seinen Daten passenden generalisierten linearen Modells. Betont wird, daß das System ›liberal‹ ist ([127], S. 166, [92], Abschnitt 5). D. h., der Benutzer kann jederzeit Vorschläge des Systems ablehnen beziehungsweise direkt GLIM-Kommandos eingeben.

Das statistische Wissen wird in Form von prädikatenlogischen Ausdrücken (genauer: HORN-Klauseln) repräsentiert. Als Implementierungssprache wurde für das Statistische Expertensystem (SIGMA-)PROLOG verwendet. Die ›numerische‹ Datenanalyse wird, wie erwähnt, mit dem in FORTRAN geschriebenen GLIM, welches über GLIMPSE aktiviert werden kann, durchgeführt.

GLIMPSE ist zur Zeit an einigen Stellen zur Erprobung installiert. Es soll ab Ende 1988 kommerziell erhältlich sein; allerdings werden für GLIMPSE Rechner mit einer Rechenleistung benötigt, die über der üblicher Arbeitsplatzrechner liegt (laut Aussage J. A. NELDER während COMPSTAT 1988). Ein in [126] zusammengestelltes Beispiel eines Benutzerdialogs mit GLIMPSE enthält Abbildung 4.

#### 4.4 PANOS

PANOS steht für *parametric and nonparametric statistics*. Entwickelt wurde das Expertensystem im Rahmen einer Dissertation von K. M. WITKOWSKI an der Universität Tübingen ([123], [124], [125]). Der Benutzer – vorausgesetzt wird eine Fachperson – soll unterstützt werden bei der Auswahl geeigneter statistischer Methoden. Es handelt sich um ein System für die Studienplanung bei konfirmatorischer Datenanalyse, das eine größere Klasse parametrischer und nichtparametrischer Tests zu berücksichtigen versucht. Das System verhält sich – im Gegensatz zu GLIMPSE – bewußt ›autoritär‹. Mit PANOS soll eine der Statistik angemessene Problembeschreibungssprache zur Verfügung gestellt werden, eine Sprache, in der von den ›rechentechnischen‹ Aspekten bestimmter Teststatistiken abstrahiert werden soll.

Die Daten werden mittels ›struktureller Information‹ beschrieben (Faktorart, Skalenniveau, ...), das statistische Wissen über Produktionsregeln. Die Methodenauswahl basiert u. a. auf der strukturellen Information über die Daten.

PANOS existiert z. Zt. als Prototypversion auf einem Arbeitsplatzrechner und ist in BASIC implementiert. PANOS ist nicht erhältlich.

#### 4.5 REX

REX steht für *Regression Expert* und wurde von W. A. GALE und D. PREGIBON in den BELL Laboratorien, Murray Hill (N.J.), entwickelt ([44], [45], [98] und insbesondere [40]). Es dient der Unterstützung von statistischen Laien, aber auch von statistischen Fachperso-

Abbildung 4. Beispieldialog in GLIMPSE (Benutzerschnittstelle, aus [126], S. 213).

```

i enter command> help task
o
  Which of the following variables should be included in
  the model ?
  1. X1
  2. X2
  3. X3
  -----
i Answer is> ?
o
  Suggested answers:
  ( X1 X2 )
  Do you wish to
  1. accept the suggested answers
  2. give your own answers
  3. receive an explanation before deciding
  -----
i Answer is> 3
o
  :
  (explanation of how this advise was derived etc.)
  :
  Do you need further explanation ?
  -----
i Answer is> no
o
  Which of the following variables should be included in
  the model ?
  1. X1                                     (1)
  2. X2
  3. X3
  -----
i Answer is> 1
o
  Suggested answers:
  ( X1 X2 )
  Do you wish to
  1. accept the suggested answers
  2. give your own answers
  3. receive an explanation before deciding
  -----
i Answer is> 1
o
  The suggested task is
  create child 1 + (X1 X2)
  Do you wish to
  1. carry out this suggestion
  2. explore the reason for this advice
  3. do something else
  -----
i Answer is >1
o executing command> create child 1 + (X1 X2)
  -----
  model selection node 2
  kernel      = (1 T X1 X2)
  free-terms = (X3 X4 X5)

```

nen bei der Anpassung (linearer) regressionsanalytischer Modelle an vorhandene Daten. REX dient, wie GLIMPSE, der explorativen Datenanalyse.

Das statistische Wissen wird vor allem mittels Produktionsregeln und Frames repräsentiert (Näheres vgl. [40], S. 201–222). Die eigentliche »numerische« Datenanalyse wird über das Statistische Auswertungssystem S durchgeführt. Das Statistische Expertensystem ist in LISP implementiert und kann S, welches in C und FORTRAN geschrieben ist, aktivieren.

Die Entwicklung an REX ist mittlerweile abgeschlossen. Das System ist erhältlich. Allerdings ist eine sehr dedizierte Hard- und Softwareumgebung erforderlich, wenn das System ohne Umprogrammierung lauffähig sein soll. Betont wird, daß die Systementwicklung einen rein experimentellen Charakter hatte und daß es sich keineswegs um ein kommerziell vertreibbares System handelt. Nach Abschluß von REX wurde in den BELL Laboratories an Nachfolgesystemen, u. a. an dem System STUDENT ([41], [42]) gearbeitet. Eine Beispielsausgabe von REX (aus [98]) befindet sich in Abbildung 5. Weitere Beispiele stehen in [40], S. 176–193.

## 5 Zur Konstruktion Statistischer Expertensysteme

### 5.1 Kritische Vorbemerkungen

Der Begriff (Statistisches) Expertensystem erscheint dem Verfasser vor allem aus zwei Gründen unglücklich gewählt zu sein: Er ist z. Zt. sehr populär, und man kann mit ihm – so scheint es zumindest – Forschungsgelder erhalten. Beide Gründe dürften für die Vielzahl von (Ankündigungen von) Expertensystemen verantwortlich sein, eine Vielzahl, die es einem schwermacht, den Weizen von der Spreu zu trennen. Die beiden Gründe wiegen schwerer als ein durch den Begriff Expertensystem verursachtes, drittes Problem: es ist teilweise schwierig – manchmal unmöglich – zu sagen, ob eine Tätigkeit nun tatsächlich eine spezifische *Expertentätigkeit* ist oder ob sie es nicht ist.

Ein Beispiel, das der Verfasser bereits in [62], S. 1, geschildert hat, soll verdeutlichen, wie bedenklich populär auch der Begriff *Statistisches Expertensystem* geworden ist:

Bei einer Vorführung eines Statistischen Auswertungssystems wurde, um die Vorzüge des Systems zu verdeutlichen, ein in der Angewandten Statistik häufig vorkommender Versuchsplan beschrieben: Es handelte sich um eine randomisierte Studie, bei der jeder Proband streng zufällig einer von zwei Behandlungen zugeteilt wurde. An jedem Probanden wurde dann – nach Behandlung – ein bestimmtes Kriterium gemessen. Ein Vorteil des Systems war nun, daß es einen Test auf Normalverteilung durchführen kann und entscheiden kann, ob anschließend »der« t-Test oder »der« WILCOXON-Test durchgeführt wird. Der Benutzer braucht nicht mehr zu wissen, ob das Kriterium normalverteilt war oder nicht. Auf die Frage, inwieweit sich die beiden Tests gegenseitig beeinflussen und, vor allem, ob das System die multiple Testsituation berücksichtigen würde, wurde dem Verfasser geantwortet, daß die derzeitige Version des Systems dies noch nicht berücksichtigte, daß aber die nächste Systemversion ein Statistisches Expertensystem wäre – womit die Frage vollständig beantwortet zu sein schien.

Hier muß zwangsläufig der Eindruck entstehen, daß mit solchen Systemen statistische Auswertungen nicht transparenter gemacht werden sollen, sondern daß im Gegenteil befürchtet werden muß, daß die nicht immer leicht zu verstehenden statistischen Methoden noch unverständlicher gemacht werden sollen.

Wie in 4.1 erwähnt, sind Arbeiten zu Expertensystemen in der Statistik teilweise älter, als es der Name »Statistisches Expertensystem« ist. Es fällt unangenehm auf, daß manche Entwickler von Statistischen Expertensystemen die in den älteren Arbeiten vorgestellten Ansätze gar nicht zu kennen scheinen.

In manchen der in letzter Zeit erschienenen Arbeiten wird außerdem der Begriff Expertensystem schlichtweg falsch verstanden und als Auswertungssystem für den statisti-

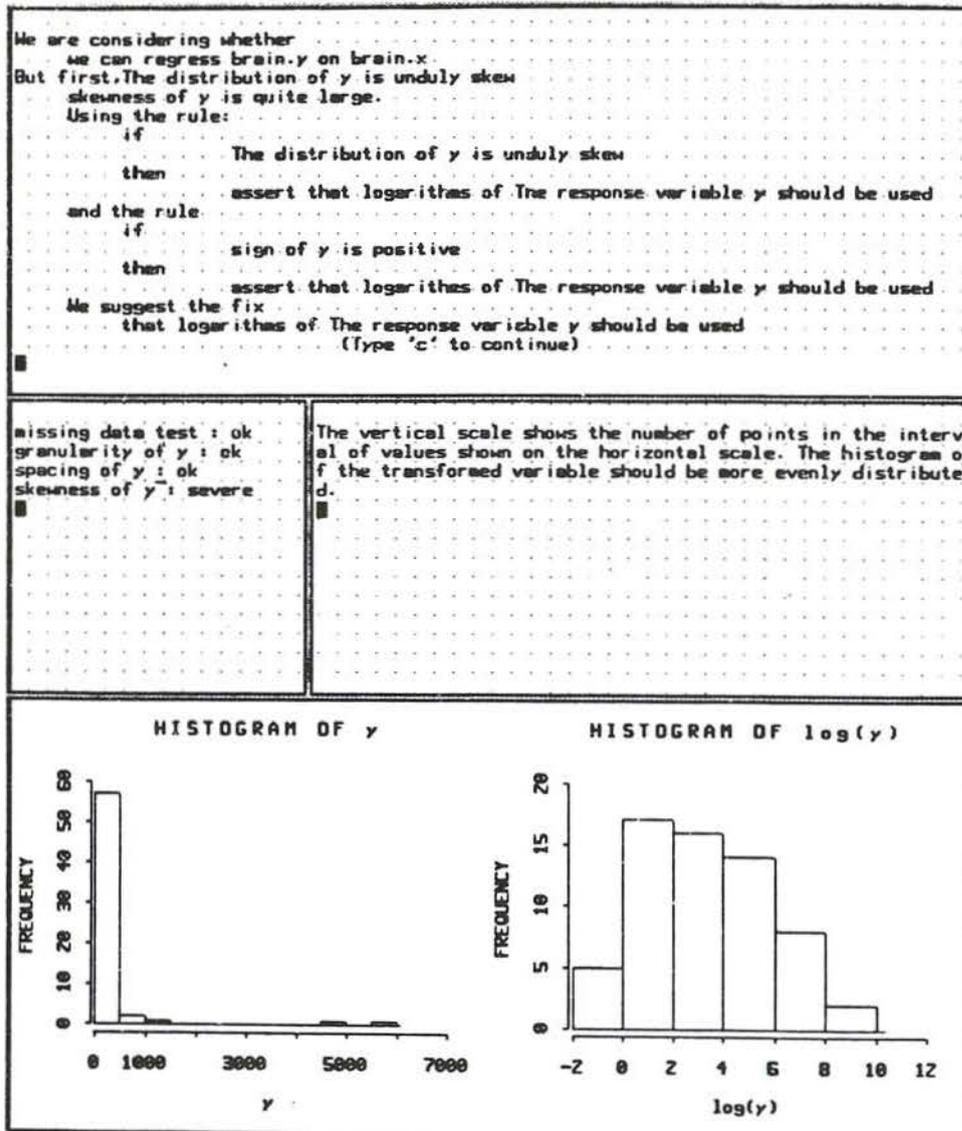


Abbildung 5. Beispielausgabe von REX (Benutzerschnittstelle, aus [98], S. 248).

schen Experten interpretiert. Diese Eigenschaft ist aber weder ein notwendiges noch ein hinreichendes Kriterium für Expertensysteme und genauso falsch wie die (manchmal vorhandene) Folgerung, jedes LISP-Programm hätte ›Künstliche Intelligenz‹ in sich (besonders wenn mit ›Maus‹ eingegeben und auf Bit-Map-Terminal ausgegeben wird) oder jedes Programm mit Produktionsregeln wäre ein Expertensystem. Auch der umgekehrte Fall, daß etwa ein in BASIC geschriebenes Programm prinzipiell kein Expertensystem sein kann, ist falsch und zeigt die Unkenntnis des Kritikers. Es muß zunächst untersucht werden, inwieweit das repräsentierte Wissen richtig ist, inwieweit die gewählte Wissensrepräsentation

tionsform beziehungsweise das gewählte Datenmodell und inwieweit der verwendete Inferenzmechanismus geeignet ist. Die Frage, ob eine imperative Sprache mit wenig Strukturierungsmöglichkeiten – wie BASIC – ein geeignetes Werkzeug ist, ist natürlich auch berechtigt, jedoch zweitrangig.

Ähnlich Kritisches ist auch zu Expertensystem-Shells anzumerken: Die Verwendung einer Expertensystem-Shell oder eines Entwicklungssystems allein löst noch kein Problem (wie manchmal angenommen zu werden scheint). Zuerst muß man das Problem erkennen, es in geeigneter Weise strukturieren und Lösungsmöglichkeiten für das Problem erarbeiten. Erst danach folgt die Auswahl geeigneter Werkzeuge und nicht umgekehrt, wie manche Hersteller von ›Shells‹ glauben machen wollen.

### 5.2 Die formale Repräsentation von Daten und Wissen und der Fehler dritter Art

Die Qualität der Problemlösungsvorschläge, die ein Statistisches Expertensystem geben kann, hängt davon ab, wieviel Informationen über das konkrete Problem – also auch über die zu erhebenden beziehungsweise auszuwertenden Daten – dem Expertensystem bekannt sind und welches Wissen über das Anwendungsgebiet in der Wissensbank enthalten ist.

Manche Statistischen Expertensysteme begnügen sich jedoch damit, einen in der Datenbank eines Statistischen Auswertungssystems bereits vorhandenen Datensatz zu analysieren. Diese Daten liegen dann häufig in codierter Form – als Ausprägungen von Variablen des Datentyps ›reelle Zahl‹ – vor. In solchen Situationen besteht offensichtlich die Gefahr, daß ein Statistisches Expertensystem aufgrund der vorhandenen, geringen Informationen über die Daten zu falschen, wenn auch in sich möglicherweise stimmigen Schlußfolgerungen über die Auswertung der Daten kommt.

Vor diesem ›Fehler dritter Art‹ – die richtige Antwort auf eine falsche Frage zu geben – hat bereits vor mehr als dreißig Jahren KIMBALL ([80]) gewarnt. Daß man in der Angewandten Statistik zunächst das Problem erkennen sollte, wird zu Recht u. a. in [76] beschrieben. Ausführliche Beispiele hierzu befinden sich in [54]. Das bloße numerische Codieren von Daten reicht nicht aus. Die negativen Konsequenzen, die hieraus resultieren, werden anhand eines Beispiels in [61] aufgezeigt.

Eine Lösungsmöglichkeit hierfür wäre die Verwendung geeigneter Daten- und Wissensrepräsentationsformen. Ein interessanter Vorschlag der für das Anwendungsgebiet Statistik zugeschnittenen, adäquaten Repräsentation von Problemen (›Daten‹) und Wissen befindet sich in [32], S. 18.

### 5.3 Datenstruktur- und Datentypkonzepte

Eine weitere Lösungsmöglichkeit ist die Verwendung geeigneter Datenstruktur- und Datentypkonzepte [61], für deren Repräsentation sich objektorientierte Sprachen eignen [96]. Gerade das Erarbeiten beispielsweise von für bestimmte medizinische Anwendungsgebiete geeigneten Datenstruktur- und Datentypkonzepten ([67]) – sei es für Statistische Auswertungssysteme allein oder in Bezug zu Statistischen Expertensystemen – und die Verbesserung der (noch immer nicht zufriedenstellenden) Integration von Datenbanksystemen und Statistischen Auswertungssystemen ([68]) bilden ja die Grundlage für die formale Repräsentation von relevantem Wissen für eine sinnvolle, interpretierbare statistische Datenanalyse.

Man kann sich manchmal des Eindrucks kaum erwehren, daß es in der Mathematischen Statistik und für einfache Anwendungsbeispiele genügt, von Realisationen reellwertiger (und möglichst normalverteilter) Zufallsvariablen auszugehen und diese Realisationen wie auch alle anderen Merkmalsausprägungen als Werte von Variablen mit Datentyp ›reellwertig‹ zu codieren und formal zu repräsentieren. Zumindest in der Angewandten Statistik benötigen wir adäquate Konzepte für die formale Repräsentation von Daten.

Daß ›ältere‹ Statistische Auswertungssysteme wie BMDP, SPSS oder auch SAS bei ihren Datenstruktur- und Datentypkonzepten immer noch Defizite aufweisen, mag verständlich

sein. Daß aber selbst bei neueren Auswertungssystemen wie DATA DESK, das hervorragende Möglichkeiten für eine interaktive statistische Datenanalyse bietet, ordinal skalierte Merkmale immer noch über reellwertige Variable codiert werden müssen (und damit dem System praktisch keine Möglichkeiten zu für den Benutzer relevanten Fehlermeldungen gegeben ist), ist bedauernd und mag zu dem Wunsch verleiten, daß vor der Entwicklung von Statistischen Expertensystemen zunächst einmal bessere Statistische Auswertungssysteme konstruiert werden sollen.

Das Erarbeiten methodischer Grundlagen für geeignete Datenstruktur- und Datentypkonzepte für Statistische Auswertungssysteme und als Voraussetzung für Statistische Expertensysteme mag nicht so populär und nicht so einfach sein wie das Aufstellen von Produktionsregeln. Es erscheint jedoch mindestens ebenso wichtig wie das Erarbeiten von Konzepten für die formale Repräsentation von Wissen. Vorschläge hierfür befinden sich in [61] und in der dort zitierten Literatur.

#### *5.4 Wissensrepräsentationsformen und Inferenzmechanismen bei regelbasierten Expertensystemen*

Das Aufstellen von Produktionsregeln – so populär es wie gesagt auch sein mag – sagt noch wenig über die Güte eines Expertensystems aus. Zur formalen Repräsentation von Wissen und Daten sollten im Abstraktionsgrad höhere und auf das Anwendungsgebiet Statistik zugeschnittene Wissensrepräsentationsformen genommen werden.

Der Produktionsregelansatz mit datenorientiertem Inferenzmechanismus ist zudem trivial, und es wäre ein Irrtum, zu glauben, daß der Stand der Expertensystementwicklung sich auf diesem Niveau befände. Man kann sich dies klarmachen, indem man sich vor Augen führt, daß eine ›Produktionsregelbasis‹, in der keine (Hintereinander-)Verkettung zwischen Regeln vorliegt (und dies scheint in manchen Expertensystemen der Fall zu sein), sich mit linearem Aufwand in die seit langem bekannte Entscheidungstabelle überführen läßt.

#### *5.5 Probabilistische Inferenzmechanismen in Statistischen Expertensystemen*

Es ist verblüffend, wie stark auch bei Statistischen Expertensystemen an den deterministischen Entscheidungstabellen gehangen wird und wie wenig beliebt – selbst unter Statistikern! – probabilistische Inferenzmechanismen ([78], [79], [82]) sind, mögen sie auf bedingten Wahrscheinlichkeiten, diskriminanzanalytischen Ansätzen oder auf anderen Ansätzen beruhen.

Auf die Eignung der wahrscheinlichkeitstheoretisch nicht ganz einfach zu interpretierenden Sicherheitsfaktoren bei regelbasierten Expertensystemen (STREITBERG in [112]: »simply pseudoscientific mumbo-jumbo«) kann der Verfasser an dieser Stelle nicht eingehen. Angemerkt sei nur, daß der zunächst auch wahrscheinlichkeitstheoretisch begründete Ansatz (vgl. z. B. [12], S. 233–262 oder [78]) später auch mißbräuchlich übernommen wurde, insbesondere indem die Sicherheitsfaktoren unmittelbar mit Wahrscheinlichkeiten gleichgesetzt wurden.

#### *5.6 Qualifikation von Benutzern und Wissensverwaltern*

Einige Statistische Expertensysteme gehen davon aus, daß auch ein statistischer Laie ein solches System benutzen kann. Dies ist nicht ganz unproblematisch. Alle der bisher existierenden Statistischen Expertensysteme decken (nach Wissen des Verfassers) nur einen kleinen Teil des Anwendungsgebiets Statistik ab. Die Problemlösungsvorschläge solcher Systeme können immer nur aus deren Teilgebiet kommen, so daß auch hier zwangsläufig die Gefahr eines Fehlers dritter Art besteht.

Diese Gefahr besteht nicht oder nur in geringem Maße, wenn Fachpersonen das System benutzen, Personen, die insbesondere so viel statistische Grundkenntnisse haben sollten, daß sie ein Problem einem Problembereich (d. h. einem Teilgebiet der Statistik) richtig

Interpretation	samba-pc	T-Test-2
<p>Der T-Test prüft, ob die Mittelwerte zweier Stichproben signifikant verschieden sind. Zu diesem Zweck wird folgende Hypothese getestet:</p> <p>Die Mittelwerte der zwei Stichproben sind gleich.</p> <p>Von unserer Bereitschaft, Fehlentscheidungen zu treffen, hängt die Folgerung über die Gültigkeit dieser Nullhypothese ab. Sind wir bereit, das Risiko zu übernehmen, im Mittel in 5 von 100 Stichproben ein Fehlurteil zu fällen, so begnügen wir uns mit einer Irrtumswahrscheinlichkeit von 0.05. Ihre individuelle Irrtumswahsch. legen Sie bitte nach Ihrem Sicherheitsbedürfnis fest.</p> <p>-RETURN- weiter im Programm</p>		

Interpretation	samba-pc	T-Test-2
Variabl.	T-Test für homogene Varianzen:	T-Test für heterogene Varianzen:
	T-Wert Freiheitsgr. Signifikanz	T-Wert Freiheitsgr. Signifikanz
FUEHRUNG	1.76 555 .080	1.73 354.65 .084
<p>Interpretation des T-Testes für homogene Varianzen:                  Der Signifikanzwert gibt die Irrtumswahrscheinlichkeit für den Fall an, daß man die Nullhypothese ablehnt.                  Oft wird ein Signifikanzniveau von 0.05 gewählt, was bedeutet, daß man in bis zu 5 von 100 Fällen einen Irrtum akzeptiert.                  Wenn wir ein Signifikanzniveau von 0.05 wählen, so ist die Nullhypothese beizubehalten,                  da der Signifikanzwert größer ist als das 5% Signifikanzniveau.</p> <p>Möchten Sie einen weiteren Test, falls berechnet, interpretiert sehen (ja: j) ?</p> <p>-RETURN- Eingabe beenden</p>		

Abbildung 6. Hilfe bei der Ergebnisinterpretation (aus [11], S. 7).

zuordnen können. Beispielsweise sollte eine statistische Fachperson ein Problem, welches mit kontingenztafelanalytischen Verfahren gelöst werden kann, als ein solches erkennen können.

5.7 Zur Qualität von Wissen

Die Qualität des in der Wissensbank repräsentierten Wissens darf ebenfalls nicht unterschätzt werden. Wissen zu repräsentieren heißt beispielsweise nicht nur, vereinfachende kochrezeptartige Entscheidungstabellen zur Auswahl statistischer Tests aus Lehrbüchern in Produktionsregelform zu bringen. Der Leser eines Lehrbuchs kann solche Kochrezepte als Hinweis nehmen, er wird jedoch (beziehungsweise sollte zumindest) die Richtigkeit dieser Hinweise in jedem Einzelfall prüfen. Ein statistisches Expertensystem ist hierzu natürlich nicht in der Lage.

Greifen wir zunächst noch einmal das in 5.1 erwähnte Beispiel auf: In Produktionsregeln ausgedrückt hieße dies (vereinfacht):

WENN Kriterium normalverteilt            DANN t-Test,  
 WENN Kriterium nicht normalverteilt    DANN WILCOXON-Test.

Dieser Hinweis zur Auswahl eines statistischen Tests allein genügt zur formalen Repräsentation von Wissen nicht. Zunächst müßte das für das Problem geeignete statistische Modell mit den dazugehörigen Test- beziehungsweise Schätzproblemen gefunden werden. Für den Fall, daß wir es dann tatsächlich mit zwei ›unabhängigen Stichproben‹ zu tun haben und daß Behandlungsunterschiede zu testen sind, ist die Entscheidung immer noch nicht so einfach, wie in den Regeln dargestellt. Ganz abgesehen davon, daß es nicht ›den‹ t-Test beziehungsweise ›den‹ WILCOXON-Test gibt, sprechen auch bei normalverteiltem Kriterium einige Argumente für den WILCOXON-Test mit Normal- beziehungsweise VAN-DER-WAERDEN-Gewichten (›Scores‹): Die asymptotisch relative PITMAN-Effizienz gegenüber der entsprechenden t-Test-Statistik ist 1, und dieser WILCOXON-Test ist auch dann ein Test zum Niveau  $\alpha$ , wenn die Normalverteilungsannahme nicht stimmt. Auf der anderen Seite, bei nicht normalverteiltem, Kriterium sprechen auch Argumente für die Zweistichproben-t-Tests. Zum Beispiel sind diese Tests asymptotisch verteilungsfrei.

Nehmen wir noch ein weiteres Beispiel: In einer Arbeit über die Entwicklung einer sogenannten Methodenbankhülle um die Arbeitsplatzrechnerversion des Statistischen Auswertungssystems SPSS, das System SAMBA/PC ([11]), werden Entscheidungs- und Interpretationshilfen für den gelegentlichen Statistikanwender präsentiert. Diese Methodenbankhülle läßt sich in vielerlei Hinsicht mit Statistischen Expertensystemen, etwa mit der ›wissensbasierten Schnittstelle‹ GLIMPSE, vergleichen.

Ein Beispiel für eine Interpretationshilfe von SAMBA/PC für den t-Test für zwei unabhängige Stichproben aus [11] befindet sich in Abbildung 6, eine Entscheidungstabelle für die Auswahl statistischer Methoden in Abbildung 7.

Obwohl der Verfasser den in [11] vorgestellten Ansatz für eine verbesserte Interpretationshilfe beziehungsweise für eine transparentere Auswahl statistischer Methoden für wichtig und im Grundsatz für richtig hält, hat er mit der Qualität des dort enthaltenen Wissens große Probleme.

So steht in Abbildung 6, daß der t-Test für das unabhängige Zweistichprobenproblem prüft, »ob die Mittelwerte zweier Stichproben signifikant verschieden sind«, als Nullhypothese wird formuliert: »Die Mittelwerte zweier Stichproben sind gleich« ([11], S. 7). Dies ist natürlich so nicht richtig und außerdem für den Benutzer verwirrend. Geprüft wird ja hier, ob sich Erwartungswerte in der Grundgesamtheit unterscheiden. Hätten die Wissensverwalter hier zunächst das statistische Modell aufgestellt und hätten sie die hier möglichen Nullhypothesen mit ihren entsprechenden Alternativhypothesen formuliert, dann wäre dieser Irrtum vermutlich gar nicht erst aufgetreten. Auch der häufige Gebrauch der Worte Signifikanz, signifikant usw. ist problematisch. Signifikanzwert läßt sich beispielsweise durch p-Wert ersetzen (und müßte natürlich jeweils für ein- beziehungsweise zweiseitige Alternativen definiert werden), die oben erwähnten ›signifikanten Unterschiede‹ entfallen sowieso.

Bei der Entscheidungstabelle in Abbildung 7 (aus [11], S. 7, vgl. auch [84], S. 536, [85], S. 306) liegt wieder das Problem der kochrezeptartigen Hinweise vor. So läßt sich beispielsweise der KRUSKAL-WALLIS-Test eben auch für intervallskalierte Kriterien, nicht nur für ordinalskalierte verwenden. Wieder würde sich vieles klären, hätte man zunächst die statistischen Modelle und die dazugehörigen Hypothesen für die in der Tabelle angegebenen statistischen Tests aufgestellt.

### 5.8 Aktive oder passive Problemlösungsunterstützung

Die Gefahr des Fehlers dritter Art ist auch dann geringer, wenn das Statistische Expertensystem nur passiv reagiert und nur den Benutzer (in einem konstruktiven Sinne!) kontrolliert und gegebenenfalls Fehler meldet ([61], S. 10). Solche Systeme – im medizinischen Anwendungsgebiet auch als kritisierende Expertensysteme bezeichnet (›expert critiquing systems‹, vgl. [64], S. 99–100) – wären, bei ausreichender Qualität des statistischen Wissens,

Unterschiedshypoth.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Zahl d. Stichproben = 1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
= 2	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0
> 2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Art der Stichprobe abh.	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0
unabh.	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1
Skalierung nom.	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
ord.	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
int.	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
CROSSTABS								X						X	
K-S-1			X												
ONEWAY												X			X
T-TEST-1						X									
T-TEST-2									X						
BINOMIAL	X														
CHISQUARE	X														
COCHRAN									X						
FRIEDMAN										X					
KENDALL											X				
K-S-2								X							
KRUSKAL														X	
MCNEMAR				X										X	
MEDIAN								X							
MOSES								X							
M-W								X							
RUNS		X													
SIGN					X										
WILCOXON					X										
W-W								X							

Abbildung 7. Entscheidungstabelle für die Auswahl statistischer Methoden (aus [11], S. 7).

von Nutzen für diejenigen Benutzer, die nicht zur statistischen Beratung kommen können beziehungsweise wollen.

In diesem Sinne argumentiert auch NAEVE in [87], wenn er wünscht, daß Statistische Expertensysteme »grobe« Fehler bei Auswertungen – er nennt als Beispiel die Analyse einer Kontingenztafel mit bestimmten Methoden der Zeitreihenanalyse – als falsch erkennen und melden kann.

5.9 Statistische Expertensysteme und Datenbanksysteme

Datenbanken mit im Routinebetrieb einer Klinik erhobenen Daten sind üblicherweise nichts anderes als (mehr oder weniger gut geplante) nichtrandomisierte Studien. Bei randomisierten Studien hat man bekanntlich den Vorteil, daß die Effekte der Störfaktoren möglichst gleichmäßig auf die einzelnen Behandlungen verteilt werden. Bei nichtrandomisierten Studien ist dies nicht der Fall. Besonders Therapievergleiche durch nichtrandomisierte Studien müssen deshalb bekanntermaßen – falls überhaupt sinnvoll – mit äußerster Sorgfalt durchgeführt werden.

Ansätze, Datenbanken mit »Routinedaten« mittels Statistischer Expertensysteme mehr oder weniger automatisch auszuwerten, werden beschrieben in [7], [8] (RX) und in [23] (MAXITAB). Die schon in [60], S. 45 und in [63], S. 314–315 und 318 beschriebene Problematik solcher Auswertungen soll auch hier nochmals aus aktuellem Anlaß anhand des von BLUM in [7] und [8] dargestellten RX-Projektes angedeutet werden.

BLUM untersuchte Möglichkeiten, eine Datenbank mit Daten von rheumakranken Patienten retrospektiv mit Hilfe eines Statistischen Expertensystems beziehungsweise eines sogenannten ›Statistikroboters‹ (»robot statistician«; [7], S. 419) zu analysieren, der anhand einer durch den Benutzer eingegebenen Fragestellung u. a. automatisch ein statistisches Modell entwerfen (»create the statistical model«; [7], S. 414), statistische Methoden auswählen (»select statistical methods«; [7], S. 414) und die Datenbestände auswerten soll (»interpret the results to determine significance« [7], S. 414).

Obwohl BLUM in den meisten Fällen von einer explorativen Analyse der Daten spricht, schreibt er aber auch in [7], S. 421: »Naturally we are interested in knowing whether a given causal relationship is statistically significant« und bringt auch unter dem Abschnitt ›medical results« in [7], S. 423 eine Tabelle mit den 15 niedrigsten p-Werten, die bei einer solchen automatischen Datenanalyse ermittelt wurden. Dort steht nicht, ob diese p-Werte als Ergebnis einer explorativen oder einer konfirmatorischen Datenanalyse interpretiert werden müssen. BLUM erwähnt zwar, daß retrospektive Auswertungen problematisch sein können, er geht jedoch fast nicht auf die Möglichkeiten von verzerrten Ergebnissen, beispielsweise aufgrund von Selektionseffekten, und gar nicht auf die Problematik des multiplen Testens ein.

Eine automatische, explorative Analyse von Daten an sich kann hilfreich sein, besonders im Sinne einer sorgfältigen Auswertung von Daten. Allerdings sind solche Auswertungen nur dann sinnvoll, wenn die Voraussetzungen für die Anwendung der benutzten statistischen Methoden beachtet werden.

Über die Problematik statistischer Auswertungen von Datenbanken mit Routinedaten, besonders bei Therapievergleichen, gibt es eine Vielzahl von Arbeiten, zum Beispiel [14], [47]. In ihnen wird zu Recht vorgeschlagen, Studien systematisch zu planen (dies läßt sich in der Medizin bis zu einem gewissen Grade auch unter Berücksichtigung von routinemäßig erhobenen Daten erreichen) und nicht Datenbanken, die für andere Zwecke aufgebaut wurden, im nachhinein unsystematisch (wenn auch möglicherweise unter Verwendung von ›Methoden der Künstlichen Intelligenz‹) auszuwerten und zu nicht interpretierbaren Ergebnissen zu gelangen.

Der Verfasser hätte auf die erneute Erwähnung des RX-Ansatzes, der ihm manchmal wie ein ›Signifikanzsuchprogramm‹ (mit nicht interpretierbaren Ergebnissen) vorkommt, aufgrund der oben erwähnten Literatur gerne verzichtet. Er hat gehofft, daß der zwar teilweise sehr interessante, aber bezüglich der Anwendung statistischer Methodik problematische RX-Ansatz bezüglich seines Statistikeils keine weitere Beachtung findet. Diese Hoffnung erfüllte sich nicht, nachdem in [43] RX als »a leading application of AI in statistics« bezeichnet wurde ([43], Teil: Folienkopien, ohne Seitenzahl).

### 5.10 Künstliche und menschliche Intelligenz in der Statistik

Von den Herausgebern der Zeitschrift wurde der Verfasser gebeten, zu der Beziehung zwischen Künstlicher Intelligenz und Statistischen Expertensystemen Stellung zu nehmen.

Der Verfasser hält sich nicht für fachlich ausreichend fundiert, um eine Meinung zu dem Fachgebiet Künstliche Intelligenz zu äußern. Er ist der Ansicht, daß – um die weitreichenden und schwerwiegenden Konsequenzen aus dem Fachgebiet zu begreifen (und um nicht nur Künstliche Intelligenz als wohlklingendes Schlagwort zu benutzen) – hierfür unter anderem umfassende philosophische, insbesondere wissenschaftstheoretische und epistemologische Kenntnisse erforderlich sind. Erstaunt sein darf man allerdings über manchen überzeugten ›KI-ler‹, nicht nur in der Statistik, der nicht mehr unterscheidet zwischen ›Wissen‹ und ›Erkenntnis‹, wenn er das englische Wort ›knowledge‹, das ja sowohl ›Wissen‹ als auch ›Erkenntnis‹ bedeuten kann, bedenkenlos mit dem deutschen Wort ›Wissen‹ gleichsetzt (als lehrreiche Lektüre sei hierzu [105] und die interessante Übersetzung von BUBSER in [106] empfohlen).

Sieht man die Entwicklung von Statistischen Expertensystemen wie von Expertensystemen allgemein unter dem Gesichtspunkt, daß die in Abschnitt 3.1 erwähnten Expertenleistungssysteme (und nicht die Expertenmodellierungssysteme) konstruiert werden sollen, dann besteht keine Beziehung zwischen Künstlicher Intelligenz und Statistischen Expertensystemen, und Ausführungen zu dem Fachgebiet Künstliche Intelligenz erübrigen sich hier.

Wir sollten neben der Entwicklung qualitativ hochwertiger Statistischer Expertensysteme auch darauf achten, daß die Ausbildung und die statistische Beratung durch den Statistiker beziehungsweise Biometriker einen qualitativ hochwertigen Standard aufweist. Schon mehrmals erwähnt wurde das Aufstellen statistischer Modelle, ohne die statistische Hypothesen nicht verständlich sind.

Entsprechendes ließe sich (Selbstkritik einbezogen) ausführen über unpräzise und oft mißverständliche Aussagen (nicht nur durch Wissensverwalter Statistischer Expertensysteme) wie ›das Ergebnis ist signifikant‹ anstelle von ›wir müssen uns für die Alternativhypothese entscheiden‹ oder ›das Ergebnis ist nicht signifikant‹ anstelle von ›wir können die Nullhypothese nicht verwerfen‹, ganz zu schweigen von hochsignifikanten oder knapp nicht signifikanten Ergebnissen. Eine präzise Wortwahl würde hier vor allem für den statistisch noch nicht sehr versierten Wissenschaftler mindestens ebenso hilfreich sein wie eine automatische statistische Methodenauswahl durch ein Statistisches Expertensystem.

Entsprechendes ließe sich auch sagen zur statistischen Planung von Studien, wenn bei der Planung zwar bewiesen wird, daß ein zu verwendender statistischer Test lokal optimal für bestimmte Alternativen ist, wenn aber nicht geprüft wird, ob geplante Doppelblindbedingungen realistisch oder ob geplante Nachsorgeuntersuchungen praktikabel sind. Auch hier ist in erster Linie der Statistiker beziehungsweise der Biometriker selbst gefordert. Ein Statistisches Expertensystem wird ihn bei solchen Aspekten, soweit sich dies zur Zeit voraussehen läßt, nicht unterstützen.

### 5.11 TURING-Tests für Statistische Expertensysteme

In [114] schlägt STREITBERG die Durchführung eines TURING-Tests für Statistische Expertensysteme vor. Dieser Vorschlag wurde in [66] (S. 69) aufgegriffen und modifiziert:

- Statistische Expertensysteme sollen anhand vorgegebener Probleme aus einem bestimmten Anwendungsgebiet geprüft werden – Probleme, für die das Statistische Expertensystem richtige Problemlösungsvorschläge machen zu können vorgibt und zu deren Problemlösung statistisches Expertenwissen notwendig ist.
- Statistische Experten auf diesem Anwendungsgebiet sollen streng zufällig zugeordnet werden (a) in die Gruppe der Experten, die ebenfalls Problemlösungsvorschläge erarbeiten und (b) in die Gruppe der Experten, die zu entscheiden haben, (1) welche Problemlösungsvorschläge nach ihrer Ansicht richtig sind und welche nicht und (2) welche Problemlösungen von dem Statistischen Expertensystem und welche von den statistischen Experten vorgeschlagen wurden.

Es versteht sich von selbst, daß ein solcher modifizierter TURING-Test jeweils in Abhängigkeit vor allem von dem vorgegebenen Anwendungsgebiet systematisch (und unter Einbeziehung statistischer Methoden) zu planen und daß er unter Blindbedingungen durchzuführen ist. Neben einer Fülle interessanter Ergebnisse ist vor allem folgendes Ergebnis wichtig: Halten die statistischen Experten aus Gruppe (b) die Problemlösungsvorschläge des Statistischen Expertensystems für ausreichend, dann müssen wir uns dafür entscheiden, daß dieses getestete System tatsächlich ein Statistisches Expertensystem für das vorgegebene Anwendungsgebiet ist.

Statistische Expertensysteme sollten durch solche modifizierten TURING-Tests geprüft werden. Diese Prüfung, verbunden mit der Offenlegung von sämtlichem in der Wissensbank enthaltenen Wissen, mit der expliziten Repräsentation des Wissens und mit der Offenlegung

des Inferenzmechanismus, wäre ein wichtiger Schritt hin zu einer besseren Transparenz von statistischen Entscheidungsprozessen.

Auch eine Übersichtsarbeit über Statistische Expertensysteme könnte dann andere Inhalte bekommen. Sie könnte Erfahrungen über den eigenen Gebrauch existierender, verfügbarer Statistischer Expertensysteme enthalten, die bei konkreten Problemen benutzt wurden. Sie brauchte nicht, wie dies hier geschehen mußte, vieles im Spekulativen zu belassen.

## 6 Schlußbemerkungen

Wir sind nun in der Lage, zu STREITBERGS Thesen Stellung zu nehmen: Zu (1) ›Es gibt keine statistischen Expertensysteme‹: Diese These ist durch die bereits existierenden Statistischen Expertensysteme widerlegt und dadurch unmittelbar auch die Thesen (2) ›Es wird auch niemals solche geben‹ und (3) ›Da es sie nicht geben kann‹. STREITBERG argumentiert allerdings mit kommerziell verfügbaren Systemen, die zum einen den Experten ersetzen sollen und die zum anderen Expertenwissen über ein so großes Teilgebiet der Statistik umfassen sollen, wie es auch ein statistischer Experte haben kann. Diese Anforderungen dürften tatsächlich nur schwer – wenn überhaupt – zu erfüllen sein. Der Philosoph BURKS vertritt die These, daß ein deterministischer endlicher Automat alle natürlichen menschlichen Funktionen ausführen kann ([13]; BURKS weist übrigens darauf hin, daß einer der Begründer der Statistik, GOTTFRIED WILHELM LEIBNIZ, ähnliche Überlegungen anstellte; vgl. [13], S. 40, 41). Letztendlich ist es diese These, mit der sich STREITBERG, in modifizierter Form, befaßt: Ein Statistisches Expertensystem kann alle wichtigen Funktionen eines statistischen Experten ausführen. Und diese These möchte er widerlegen.

Bei den praktischen Arbeiten im Zusammenhang mit Statistischen Expertensystemen spielt diese These jedoch (außer möglicherweise bei der Vermarktung einiger Systeme), wie in den vorhergehenden Abschnitten ausgeführt, keine Rolle. Es geht vielmehr darum, daß die Benutzer eines Statistischen Expertensystems durch ein solches System unterstützt werden, sei es beispielsweise dadurch, daß die u. a. von NELDER und CHAMBERS erwähnte Gefahr falscher Auswertungen reduziert werden kann oder daß die Benutzer bei der (explorativen) Datenanalyse durch Statistische Expertensysteme besser als bisher unterstützt werden.

Die Inferenzmechanismen solcher Systeme sind jedoch erfahrungsgemäß – bezogen auf deren Zeitkomplexität – relativ unproblematisch: Sie liegen größtenteils in den Komplexitätsklassen  $P$  oder  $NP$ , also in der Klasse der Algorithmen, die in polynomialer Zeit auf einer deterministischen oder einer nichtdeterministischen TURING-Maschine ablaufen. Die von STREITBERG vorgebrachten komplexitätstheoretischen Argumente greifen für solche vergleichsweise ›einfachen‹ Expertensysteme noch nicht.

Es geht bei der Konstruktion Statistischer Expertensysteme aber auch darum, daß dabei – sozusagen als erwünschter Nebeneffekt – der Statistiker als Wissensverwalter gezwungen wird, sein Wissen formal zu repräsentieren und geeignete Inferenzmechanismen zum Erarbeiten von Problemlösungen anhand dieses Wissens vorzuschlagen. Dadurch können wir eine bessere Transparenz von Entscheidungsprozessen erreichen. Für IMMICH zählen Reproduzierbarkeit und das Erkennen notwendiger und genügender Bedingungen zu den Zielen klinischer Forschung ([73], S. 3). Die Statistik, strebt danach, eine exakte Wissenschaft zu sein, und dies nicht nur in der Mathematischen Statistik, sondern auch in möglichst vielen Bereichen der Angewandten Statistik. Die Entwicklung Statistischer Expertensysteme kann bei diesem Streben nach mehr Transparenz, und gerade auch nach der von IMMICH erwähnten Reproduzierbarkeit und nach dem Erkennen notwendiger und genügender Bedingungen für die Entscheidungsprozesse in der Statistik, einen wichtigen Beitrag leisten. In diesem Sinne ist der Verfasser derselben Ansicht wie STREITBERG, wenn er in

These (4) behauptet: ›Dennoch ist es sinnvoll, an den Problemstellungen derartiger Systeme zu arbeiten‹. STREITBERG ergänzt richtig, daß ein bloßes Anhäufen inhaltlich nicht besonders relevanter Regeln noch keine Lösung darstellt. Entsprechendes gilt für ›wissensbasierte Signifikanzsuchprogramme‹, unabhängig davon, ob sie die ›Signifikanz‹ in Datenbanken suchen oder anderswo. Nicht außer acht gelassen werden sollte auch die von ZELEN ([128]) befürchtete Gefahr der Schematisierung von Entscheidungsprozessen in der Statistik durch Statistische Expertensysteme.

Das Fachgebiet *Computational Statistics* (oder deutsch ›Informatik in der Statistik‹) befaßt sich mit dem Einsatz und mit der Untersuchung von Methoden und Werkzeugen der Informatik für die Statistik (in Anlehnung an [118]). In dieses Fachgebiet gehört auch die Entwicklung und Bewertung Statistischer Expertensysteme. Daß diese Entwicklung nicht einfach ist und daß hier nichttriviale Probleme zu lösen sind, dürfte anhand dieser Ausarbeitung deutlich geworden sein. Damit die Entwicklung Statistischer Expertensysteme zum Nutzen der Statistik und der Biometrie verläuft, müssen wir geeignete Methoden der Statistik und der Informatik sowie geeignete Werkzeuge der Informatik verwenden. Es ist eine wichtige Aufgabe der entsprechenden Arbeitsgruppen in den wissenschaftlichen Fachgesellschaften, die Entwicklung Statistischer Expertensysteme zu fördern und gleichzeitig darauf zu achten, daß diese Systeme dem derzeitigen Stand der Entwicklung von Statistik und Informatik gerecht werden.

### Danksagung

Der Verfasser ist J. A. NELDER, London, B. STREITBERG, Hamburg, und K. M. WITKOWSKI, Tübingen, für deren Kritik und deren Anmerkungen zu Vorversionen dieser Arbeit sehr zu Dank verpflichtet.

### Literaturverzeichnis

1. ADAMS, I. D., M. CHAN, P. C. CLIFFORD, W. M. COOKE, V. DALLOS, F. T. DE DOMBAL, M. H. EDWARDS, D. M. HANCOCK, D. J. HEWETT, N. MCINTYRE, P. G. SOMERVILLE., D. J. SPIEGELHALTER, J. WELLWOOD, D. H. WILSON (1986). Computer Aided Diagnosis of Acute Abdominal Pain: A Multicentre Study. *Br. Med. J.* **293**, 800–804.
2. BARRA, R., F. BRODEAU, G. ROMIER, B. VAN CUTSEN (Hrsg.) (1977): *Recent Developments in Statistics*. Amsterdam: North Holland.
3. BECKER, R. A., J. M. CHAMBERS (1984). *S: An Interactive Environment for Data Analysis and Graphics*. Belmont, Ca.: Wadsworth.
4. BERZUINI, C., G. ROSS, C. LARIZZA (1986). Developing Intelligent Software for Non-Linear Model Fitting as an Expert System. In: [28], 259–264.
5. BIBEL, W., N. EISINGER, J. SCHNEEBERGER, J. SIEKMANN (Hrsg.) (1987). *Studien- und Forschungsführer Künstliche Intelligenz*. Berlin: Springer.
6. BLASER, A., C. HACKL (Hrsg.) (1977). *Interactive Systems*. Berlin: Springer.
7. BLUM, R. L. (1982). Discovery, Confirmation, and Incorporation of Causal Relationships from a Large Time-Oriented Clinical Database: The RX project. *Comp. and Biom. Res.* **15**, 164–187; ebenfalls erschienen in [24], 399–425 und von dort zitiert.
8. BLUM, R. L. (1982). Discovery and Representation of Causal Relationships from a Large Time-Oriented Clinical Database: The RX project. Berlin: Springer.
9. BOARDMAN, T. J. (Hrsg.) (1986). *Computer Science and Statistics: Proc. of the 18th Symposium on the Interface*. Washington, D.C.: Am. Statist. Assoc.
10. BODENDORF, H. (1981). Unterstützung der Anwender von Statistiksoftware – Ein Methodenbankrahmen um das Programmpaket SPSS. *Reer.-pol.-Dissertation*, Universität Erlangen-Nürnberg.
11. BODENDORF, F., U. OSIANDER (1988). Eine Methodenbankhülle um das Programmpaket SPSS/PC+ für Gelegenheitsanwender. *Ang. Informatik* **1/88**, 1–8.

12. BUCHANAN, B. G., E. H. SHORTLIFFE (Hrsg.) (1984). *Rule-based Expert Systems – The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, Ma.: Addison-Wesley.
13. BURKS, A. W. (1972/73). Logic, Computers, and Men. *Proc. and Addresses of the Am. Philosoph. Assoc.* **46**, 39–57.
14. BYAR, D. P. (1980). Why Data Bases Should not Replace Randomized Clinical Trials. *Biometrics* **36**, 337–342.
15. CABLE, D., B. ROWE (1987). *Software for Statistical and Survey Analysis*. London: Study Group on Computers in Survey Analysis, British Informatics Society.
16. CARLSEN, F., I. HEUCH (1986). EXPRESS – An Expert System Utilizing Standard Statistical Packages. In: [28], 265–270.
17. CAUSSINUS, H., P. ETtingER, R. TOMASSONE (Hrsg.) (1982). *COMPSTAT 1982*. Wien: Physika.
18. CHAMBERS, J. M. (1981): Some Thoughts on Expert Software. In: [30], 36–40.
19. CHAMBERS, J. M. (1983): The New Future of Data Analysis. *Bulletin of the International Statistical Institute, Proc. 44th Session, L*, Book 1, 97–103.
20. CHAMBERS, J. M., W. A. GALE, D. PREGIBON (1988). On the Existence of Expert Systems. *Stat. Software Newsl.* **14**, 63–66.
21. CHAMBERS, J. M., D. PREGIBON, E. R. ZAYAS (1981). Expert Software for Data Analysis – An Initial Experiment. *Proc. 43rd Session of the ISI, XLIX, Book 1*, 294–303.
22. CHARNIAK, E., D. McDERMOTT (1985). *Introduction to Artificial Intelligence*. Reading, Ma.: Addison-Wesley.
23. CHOWDHURY, S. I., O. B. WIGERTZ (1988). Extraction of Knowledge from a Database. In: [104], 338–342.
24. CLANCEY, W. J., E. H. SHORTLIFFE (Hrsg.) (1984). *Readings in Medical Artificial Intelligence – The First Decade*. Reading, Ma.: Addison-Wesley.
25. CORSTEN, L. C. A., J. HERMANS (Hrsg.) (1978). *COMPSTAT 1978*. Wien: Physika.
26. DAMBROISE, E., P. MASSOTTE (1986). MUSE: An Expert System in Statistics. In: [28], 271–276.
27. DARIUS, P. L. (1986). Building Expert Systems with the Help of Existing Statistical Software: An Example. In: [28], 277–282.
28. DE ANTONI, F., N. LAURO, A. RIZZI (Hrsg.) (1986). *COMPSTAT 1986*. Heidelberg: Physika.
29. DORDA, W., B. SCHNEIDER, K. A. FRÖSCHL, W. GROSSMANN (1988). Ein statistisches Expertensystem-Modell am Beispiel deskriptiver Statistik. *Österr. Z. f. Statistik u. Informatik* **18**, 172–180.
30. EDDY, F. (Hrsg.) (1981). *Computer Science and Statistics: Proc. of the 13th Symposium on the Interface*. New York: Springer.
31. EDWARDS, D., N. E. RAUN (Hrsg.) (1988). *COMPSTAT 1988*. Heidelberg: Physika.
32. ELLIMAN, A., K. M. WITKOWSKI (1987). The Impact of Expert Systems on Statistical Database Management. *Stat. Software Newsl.* **13**, 14–18.
33. FAULBAUM F. (Hrsg.) (1989). *Fortschritte der Statistik-Software 2: 5. Konferenz über die wissenschaftliche Anwendung von Statistik-Software, Heidelberg, 1989*. Stuttgart: Fischer (im Druck).
34. FAULBAUM, F., H.-M. UEHLINGER (Hrsg.) (1988). *Fortschritte der Statistik-Software 1: 4. Konferenz über die wissenschaftliche Anwendung von Statistik-Software, Heidelberg, 1987*. Stuttgart: Fischer.
35. FRANCIS, I. (1981). *Statistical Software: A Comparative Review*. New York: North Holland.
36. FRÖSCHL, K. A. (1984). Expertensysteme in der statistischen Datenverarbeitung. *Mitteilungsblatt d. Österr. Statist. Ges.* **14**, 167–179.
37. FRÖSCHL, K. A., W. GROSSMANN (1986). Knowledge Base Supported Analysis of Longitudinal Data. In: [28], 289–294.
38. FRÖSCHL, K. A., W. GROSSMANN (1988). Statistical Structures for Analyzing Time-dependent Observations. In [46], 145–160.
39. GALE, W. A. (Hrsg.) (1986). *Artificial Intelligence and Statistics*. Reading, Ma.: Addison-Wesley.
40. GALE, W. A. (1986). REX Review. In: [39], 173–227.
41. GALE, W. A. (1986). Student Phase 1 – A Report on Work in Progress. In: [39], 239–265.
42. GALE, W. A. (1987). STUDENT: A Tool for Constructing Consultation Systems for Data Analysis. *Bulletin of the International Statistical Institute, Proc. 46th Session, LII, Book 3*, 305–321.

43. GALE, W. A. (1988). Statistical Applications of Artificial Intelligence and Knowledge Engineering. Bericht No. 311404-2399, File Case 20878 (verteilt bei COMPSTAT 1988).
44. GALE, W. A., D. PREGIBON (1982). An Expert System for Regression Analysis. In: [72], 110-117.
45. GALE, W. A., D. PREGIBON (1984). Constructing an Expert System for Data Analysis by Working Examples. In: [69], 227-236.
46. GAUL, W., M. SCHADER (Hrsg.) (1988). Data, Expert Knowledge and Decisions. Berlin: Springer.
47. GREEN, S. B., D. P. BYAR (1984). Using Observational Data from Registries to Compare Treatments: The Fallacy of Omnimetrics. Stat. in Medicine 3, 361-370.
48. GROSSMANN, W., K. A. FRÖSCHL (1988). Formale Wissens- und Verfahrensmodellierung in der statistischen Datenverarbeitung. In: [74], 71-85.
49. HAHN, G. J. (1985). More Intelligent Statistical Software and Statistical Expert Systems: Future Directions. Am. Statistician 39, 1-8.
50. HAJEK, P., J. INVANEK (1982). Artificial Intelligence and Data Analysis. In: [17], 54-60.
51. HAND, D. J. (1984). Statistical Expert Systems: Design. The Statistician 33, 351-369.
52. HAND, D. J. (1985). Statistical Expert Systems: Necessary Attributes. J. Applied Statistics 12, 19-23.
53. HAND, D. J. (1986). Expert Systems in Statistics. Knowledge Engineering Rev. 1, 2-10.
54. HAND, D. J. (1986). Patterns in Statistical Strategy. In: [39], 335-387.
55. HAND, D. J. (1987). A Directory of Statistical Expert System Research. Erhältlich von: D. J. HAND, Biometrics Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, Großbritannien.
56. HAND, D. J. (1987). A Statistical Knowledge Enhancement System. J. R. Statist. Soc. A 150, 334-345.
57. HARMON, P., D. KING (1987). Expertensysteme in der Praxis. München: Oldenbourg.
58. HAUX, R. (1983). How to Detect and Prevent Errors in Computer-Supported Statistical Analysis: An Example. Meth. Inform. Med. 22, 87-92.
59. HAUX, R. (1983/84). Statistical Analysis Systems - Construction and Aspects of Method Design. Stat. Software Newsl. 9, 106-115 und 10, 14-27.
60. HAUX, R. (1985). Datenbankaspekte bei Statistischen Auswertungssystemen. EDV in Med. u. Biol. 16, 41-46.
61. HAUX, R. (1986). Database Management and Statistical Data Analysis: The Need for Appropriate Data Types and Data Structure Types. Proc. XIIIth Int. Biometric Conference, Invited Paper Session 14, 1-15. Seattle, Wa.: Univ. of Washington.
62. HAUX, R. (Hrsg.) (1986). Expert Systems in Statistics. Stuttgart: Fischer.
63. HAUX, R. (1986). Statistical Expert Systems - Some Problems and Some New Views. In: [111], 313-322.
64. HAUX, R. (1987). Expertensysteme in der Medizin. Habilitationsschrift, RWTH Aachen.
65. HAUX, R. (1987/88). Expertensysteme in der Medizin - Eine einführende Übersicht. Softw. Kurier 1, 65-77 und 2, 1-11.
66. HAUX, R. (1988). On the Existence of Expert Systems. Stat. Software Newsl. 14, 68-69.
67. HAUX, R., F. GERNETH (1989). Non-Standard Data Models for Statistical Data Analysis. Erscheint in: Bulletin of the International Statistical Institute, Proc. 47th Session, LIII.
68. HAUX, R., K.-H. JÖCKEL (1986). Database Management and Statistical Data Analysis: The Need for Integration and for Becoming More Intelligent. In: [28], 407-414.
69. HAVRANEK, T., Z. SIDAK, M. NOVAK (Hrsg.) (1984). COMPSTAT 1984. Wien: Physika.
70. HAYES-ROTH, F., D. A. WATERMAN, D. B. LENAT (Hrsg.) (1983). Building Expert Systems. Reading, Ma.: Addison-Wesley.
71. HECKERMAN, D. (1986). Probabilistic Interpretations for MYCIN's Certainty Factors. In: [78], 167-196.
72. HEINER, K. W., R. S. SACHER, J. W. WILKINSON (Hrsg.) (1982). Computer Science and Statistics: Proc. of the 14th Symposium on the Interface. New York: Springer.
73. IMMICH, H. (1983). Methodische Voraussetzungen klinischer Forschung. MedWelt 34 (14), 3-6.
74. JANKO, W. (Hrsg.) (1988). Statistik, Informatik und Ökonomie. Berlin: Springer.
75. JESDINSKY, H. J. (1972). Diagnose-Modelle in der Medizin. Meth. Inform. Med. 11, 48-59.
76. JÖCKEL, K.-H. (1986). Statistical Expert Systems and the Statistical Consultant - Considerations about the Planning Stage of Clinical Studies. In: [62], 27-43.
77. JONES, B. (1980). The Computer as a Statistical Consultant. Bias 7, 168-194.

78. KANAL, L. N., J. F. LEMMER (Hrsg.) (1986). *Uncertainty in Artificial Intelligence*. Amsterdam: North Holland.
79. KANAL, L. N., J. F. LEMMER (Hrsg.) (1988). *Uncertainty in Artificial Intelligence 2*. Amsterdam: North Holland.
80. KIMBALL, A. W. (1957). Errors of the Third Kind in Statistical Consulting. *J. Am. Stat. Assoc.* **52**, 133–142.
81. LAMINGER, B., W. DORDA, C. REICHETZEDER (1986). WAMASTAT – Ein System zur Benutzerführung bei der Auswertung klinischer Studien. *EDV in Med. u. Biol.* **17**, 1–5.
82. LAURITZEN, S. L., D. J. SPIEGELHALTER (1988). Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (mit Diskussion). *J. R. Statist. Soc. B.* **50**, 157–224.
83. Lektorat des B.I.-Wissenschaftsverlags (Hrsg.) (1988). *Duden »Informatik«*. Mannheim: Dudenverlag.
84. MERTENS, P., F. BODENDORF (1979). Interaktiv nutzbare Methodenbanken – Entwurfskriterien und Stand der Verwirklichung. *Ang. Informatik* **12/79**, 533–541.
85. MERTENS, P., W. NEUWIRTH, W. SCHMITT (1977). Verknüpfung von Daten- und Methodenbanken, dargestellt am Beispiel der Analyse von Marktwirtschaftsdaten. In: [6], 273–313.
86. MOLENAAR, I. W. (1988). Statistical Consultants and Statistical Expert Systems. In: [31], 187–192.
87. NAEVE, P. (1986). Epilogue. In: [62], 182–184.
88. NELDER, J. A. (1977). Intelligent Programs, The Next Stage in Statistical Computing. In: [2], 79–86.
89. NELDER, J. A. (1978). The Future of Statistical Software. In: [25], 11–19.
90. NELDER, J. A. (1987). AI and Generalized Linear Modeling: An Expert System for GLIM. In: [97], 36–44.
91. NELDER, J. A. (1988). How Should the Statistical Expert System and its User See Each Other? In: [31], 107–116.
92. NELDER, J. A. (1988). The Role of Expert Systems in Statistics. In: [34], 175–182.
93. NELDER, J. A., D. E. WOLSTENHOLME (1986). A Front-End for GLIM. In: [9], 113–117.
94. NILSSON, N. J. (1982). *Principles of Artificial Intelligence*. Berlin: Springer.
95. OLDFORD, R. W., S. C. PETERS (1984). Building a Statistical Knowledge Based System with MINIMYCIN. *Proc. Am. Statist. Assoc. (statistical computing section)*.
96. OLDFORD, R. W., S. C. PETERS (1986). Object-Oriented Data Representations for Statistical Data Analysis. In: [28], 301–306.
97. PHELPS, B. (Hrsg.) (1987). *Interactions in Artificial Intelligence and Statistical Methods*. Aldershot, England: Gower.
98. PREGIBON, D., W. A. GALE (1984). REX: An Expert System for Regression Analysis. In: [69], 242–248.
99. PUPPE, F. (1987). *Diagnostisches Problemlösen mit Expertensystemen*. Berlin: Springer.
100. PUPPE, F. (1988). *Einführung in Expertensysteme*. Berlin: Springer.
101. RASCH, D., V. GUIARD, G. NÜRNBERG, E. RUDOLPH, F. TEUSCHER (1987). The Expert System CADEMO. *Stat. Software Newsl.* **13**, 107–114.
102. RASCH, D., G. NÜRNBERG, K. BUSCH (1988). CADEMO – Ein Expertensystem zur Versuchsplanung. In: [34], 193–201.
103. RICHTER, M. M. (1986). Architecture and Applications of Expert Systems. In: [62], 9–24.
104. RIENHOFF, U., U. PICCOLO, B. SCHNEIDER (Hrsg.) (1988). *Expert Systems and Decision Support in Medicine*. Berlin: Springer.
105. RUSSEL, B. (1912). *The Problems of Philosophy*. Elfte Neuauflage 1967: Oxford: University Press.
106. RUSSEL, B. (1967). *Probleme der Philosophie*. Aus dem Englischen übersetzt und mit einem Nachwort versehen von E. BUBSER, Frankfurt/M.: Suhrkamp.
107. SCHNEIDER, H.-J. (Hrsg.) (1986). *Lexikon der Informatik und Datenverarbeitung*. München: Oldenbourg.
108. SCHNUPP, P., U. LEIBRANDT (1986). *Expertensysteme – Nicht nur für Informatiker*. Berlin: Springer.
109. SCHNUPP, P., C. T. NGUYEN HUU (1987). *Expertensystem-Praktikum*. Berlin: Springer.
110. SHORTLIFFE, E. H., B. G. BUCHANAN, E. A. FEIGENBAUM (1979). Knowledge Engineering for Medical Decision Making: A Review of Computer-Based Clinical Decision Aids. *Proc. IEEE* **67**: 1207–1224.

111. STOYAN, H. (Hrsg.) (1986). GWAI-85, 9th German Workshop on Artificial Intelligence. Berlin: Springer.
112. STREITBERG, B. (1986). Epilogue. In: [62], 184–185.
113. STREITBERG, B. (1987). Expertensysteme in der Statistik – Experten oder Ignoranten? EDV in Med. u. Biol. **18**, 90–94; ebenfalls erschienen in [34], 183–192.
114. STREITBERG, B. (1988). On the Nonexistence of Statistical Expert Systems – Critical Remarks on Artificial Intelligence in Statistics (mit Diskussion). Stat. Software Newsl. **14**, 55–74.
115. STREITBERG, B., P. NAEVE (1986). A Modestly Intelligent System for Identification, Estimation, and Forecasting of Univariate Time Series: A<sup>1</sup>: ARIMA, Artificial Intelligence, and APL2. In: [62], 111–137.
116. TRAMPISCH, H. J. (1982). Zuordnungsprobleme in der Medizin. Berlin: Springer.
117. VELLEMAN, P. F., A. Y. VELLEMAN (1988). Data Desk Handbook. Northbrook, IL: Odesta Corporation.
118. VICTOR, N. (1984). Computational Statistics – Tool or Science? Stat. Software Newsl. **10**, 105–116.
119. WATERMAN, D. A. (1985). A Guide to Expert Systems. Reading, Ma.: Addison-Wesley.
120. WEISS, S. M., C. A. KULIKOWSKI (1984). A Practical Guide to Designing Expert Systems. London: Chapman and Hall.
121. WESTERHOFF, T., P. NAEVE (1988). On Inference Process. In: [31], 107–116.
122. WINSTON, P. H. (1984). Artificial Intelligence, zweite Auflage. Reading, Ma.: Addison-Wesley.
123. WITKOWSKI, K. M. (1985). Ein Expertensystem zur Datenhaltung und Methodenauswahl für statistische Auswertungen. *Ref.-nat.-Dissertation*, Universität Stuttgart.
124. WITKOWSKI, K. M. (1986). An Expert System for Testing Statistical Hypotheses. In: [9], 438–443.
125. WITKOWSKI, K. M. (1986). Estimating and Testing Statistical Hypotheses: Strategies for Knowledge Engineering. In: [62], 139–154.
126. WITKOWSKI, K. M. (1988). Dialogformen intelligenter Benutzerschnittstellen für statistische Auswertungen. In: [34], 212–225.
127. WOLSTENHOLME, D. E., J. A. NELDER (1986). A Front End for GLIM. In: [62], 155–177.
128. ZELEN, M. (1983). Biostatistical Science as a Discipline – A Look into the Future. *Biometrics* **39**, 827–837.

Anschrift des Verfassers: Prof. Dr. Reinhold Haux, Eberhard-Karls-Universität Tübingen, Institut für Medizinische Informationsverarbeitung, Westbahnhofstr. 55, D-7400 Tübingen

## Kommentare

### W. G. Dorda

An der Universität Wien beschäftigt sich seit einigen Jahren eine Arbeitsgruppe mit der konkreten Entwicklung eines »Statistischen Expertensystems« (1). Einige der dabei aufgetretenen Probleme werden kurz skizziert.

#### Bedarf an ausformulierten statistischen Auswertstrategien

Die zentralen Probleme bei der Entwicklung Statistischer Expertensysteme (SES) liegen im Bereich der Statistik. Die Entscheidungsregeln, welche in der angewandten Statistik zur Wahl eines bestimmten Verfahrens führen, sind nur selten exakt definiert und unterscheiden sich zudem manchmal zwischen den »Statistischen Schulen«. Als Beispiele seien graphische Darstellungen (etwa der Verteilung einer Variablen oder des Zusammenhangs

zwischen Variablen) genannt: Die Schlüsse, welche ein erfahrener Statistiker aus der Betrachtung solcher Graphiken zieht, lassen sich nur schwer formalisieren. Die Definition entsprechender Heuristiken und die Entwicklung fundierter Auswertstrategien sind daher eine positive Herausforderung an die statistische Grundlagenforschung, welcher sich die Statistiker unbedingt stellen sollten: Da der Bedarf an statistischer Anleitung durch Softwaresysteme zweifellos vorhanden ist, werden solche Systeme sonst auch ohne die entsprechenden theoretischen Grundlagen aus kommerziellen Gründen angeboten werden.

### Gefahr der Suggestion falscher Sicherheit

Die allgemeine Verfügbarkeit statistischer Auswertsysteme führt oft zum inadäquaten Einsatz statistischer Methoden. Obwohl SES meist entwickelt werden, um diese Situation zu verbessern, könnte durch die Suggestion falscher Sicherheit genau das Gegenteil eintreten. SES dürfen dem Benutzer also auf keinen Fall ein unkritisches Sicherheitsgefühl vermitteln, sondern ihn vor allem auf Gefahren falscher Statistikanwendung aufmerksam machen. Angesichts der unzähligen Statistiken, welche tagtäglich laienhaft durchgeführt werden, wäre die Bedeutung solcher »kontrollierender Systeme« kaum zu überschätzen. In diesem Zusammenhang ist es bedauerlich, daß bei den meisten der zur Verfügung stehenden statistischen Auswertsystemen die einzige kritische Anmerkung die Meldung »WARNING: 30 % of the cells have expected counts less than 5. Chi-square may not be a valid test« ist (4, Kontingenztafelanalyse) ist. Der Verweis auf ein so banales Beispiel mag bei der Besprechung »Statistischer Expertensysteme« anachronistisch wirken, aber in der Praxis werden doch fast alle statistischen Analysen mit solchen Auswertsystemen durchgeführt. Der Einbau von Kontrollmechanismen in die statistischen Auswertsysteme könnte daher viel zur Verhinderung des falschen Einsatzes statistischer Verfahren beitragen.

Solche Systeme werden von den Anwendern aber nur dann akzeptiert werden, wenn sie nicht nur als Bremse, sondern auch als *positive Führung* in Erscheinung treten. Systeme, welche sich nur kritisch bemerkbar machen, könnten von den Anwendern als zu unbequem empfunden werden und unbenutzt bleiben. In der Praxis ist die Balance zwischen der »frustrierenden Bremse« und einem »Sicherheit vorgaukelnden Expertensystem« nicht leicht zu erzielen. Wir sehen einen praktikablen Weg in Richtung eines »semi-autoritären« SES, dessen dezidiert gegebene Warnungen oder Vorschläge der Benutzer ablehnen kann. Das Ziel sollte ja sein, den Benutzer zur kritischen Eigenverantwortlichkeit anzuleiten, wenn eine anerkannte und objektive Lösung des Problems nicht möglich ist.

Welche Bedeutung der präzisen und didaktisch adäquaten Formulierung im Rahmen des Bildschirmdialoges zukommt, zeigen auch die kritischen Bemerkungen von HAUX (3, Kapitel 5.7 und 5.10) über falsche oder irreführende Formulierungen. Mißverständliche Anweisungen könnten in vielen Fällen durch die Angabe einfacher Beispiele verhindert werden.

Die Frage, ob diese aus dem konkreten Anwendungsgebiet sein sollten, führt zu Überlegungen, ob ein »Statistisches Expertensystem« auch *spezifisches Wissen aus dem Anwendungsfach* benötigt. Dem Vorteil der besseren und eindeutigeren Kommunikation mit dem Benutzer steht der Nachteil einer Einschränkung des Benutzerkreises bzw. ein jeweils fachspezifischer Adaptierungsaufwand gegenüber. Bei der konkreten Formulierung des Bildschirmdialoges und der Auswahl konkreter Beispiele müssen selbstverständlich auch Überlegungen über die Qualifikation der Benutzer auf dem Gebiet der Statistik (vgl. 3, Kapitel 2.3 und 3.3) berücksichtigt werden.

Auf jeden Fall sollte bei der Entwicklung von SES dem *didaktischen Aspekt* größte Beachtung geschenkt werden. Diese kommen so in positive Wechselwirkung mit Systemen der »Computerunterstützten Lehre« (wie dies auch bei Medizinischen Expertensystemen der Fall ist, vgl. z. B. (2)).

### Heterogene Anforderungen an ein SES

Die Anforderungen an ein SES umfassen ein sehr großes Spektrum: Ein *SES zur konfirmatorischen Datenanalyse* sollte bereits die Studienplanung mit der Modelldefinition umfassen und auch alle folgenden Einzelschritte unterstützen, also zum Beispiel: Aufstellen des statistischen Modells, Festlegen der Hypothesen und der Analysemethodik, Bestimmen der Stichprobengrößen, Randomisierung, Formularentwurf, Überwachung der Datenerhebung, Kontrolle eventueller Zwischenauswertungen (etwa zum Erkennen eines notwendigen Studienabbruches), Auswertung unter Berücksichtigung der multiplen Testproblematik, Ergebnisinterpretation u. v. a. Das Spektrum der Informatikmethoden zur Unterstützung dieser einzelnen Phasen ist sehr groß; man denke etwa an die Unterschiede zwischen den Komponenten zur Unterstützung der Randomisierung und denen zur Ergebnisinterpretation.

Bei einem *SES zur explorativen Datenanalyse* werden einige dieser Punkte weniger Bedeutung haben, dafür wird sich ein Schwerpunkt bei den entsprechenden – graphisch oft sehr aufwendigen – Methoden ergeben. Weiter werden sehr viele Entscheidungen über die weitere Auswertstrategie aufgrund der Datensituation gefällt werden. Dies führt weg von den »Statistischen Konsultationssystemen« zu den »Hypothesen generierenden Systemen«, welche vollautomatisch aus Daten Hypothesen ableiten (und »beweisen«) wollen.

### Zusammenfassung

Die Entwicklung »Statistischer Expertensysteme« liefert interessante Anstöße zur *Entwicklung methodisch fundierter Auswertstrategien*. Das zumeist angestrebte Ziel, die *Qualität des Statistikeinsatzes in der Praxis zu verbessern*, ist gegenwärtig sicherlich *nur auf ganz kleinen Teilgebieten der Statistik zu erreichen*. Durch den Einsatz nicht ganz so »unintelligenter« Softwaresysteme wie die derzeit gebräuchlichen statistischen Auswertssysteme ist aber eine langsame und schrittweise Verbesserung zu erhoffen, welche von den Anwendern sicherlich sehr begrüßt werden wird. Dabei sollte man sich nicht scheuen, auch weniger avantgardistische Methoden der Informatik einzusetzen, solange diese zum Ziel führen. Die Frage, ob solche Systeme als »Statistische Expertensysteme« zu bezeichnen sind, ist letztlich sekundär; zur Vermeidung eines *falschen Sicherheitsgefühls* bei den Anwendern sollte man diesen Begriff ohnehin eher meiden und jeder praktischen Applikation einen entsprechenden Hinweis über die Grenzen des SES voranschicken.

### Literatur

1. DORDA, W., B. SCHNEIDER, K. FRÖSCHL, W. GROSSMANN: WAMASTEX: Ein statistisches Expertensystem-Modell am Beispiel deskriptiver Statistik. Österreichische Zeitschrift für Statistik und Informatik **3**, 172–180 (1988).
2. FIRST, M. B., L. J. SOFFER, R. A. MILLER (1985): QUICK (Quick Index to Caduceus Knowledge): Using the Internist-1/Caduceus Knowledge Base as an Electron Textbook of Medicine. Computers and Biomedical Research **18**, 137–165.
3. HAUX, R.: Statistische Expertensysteme. Biometrie und Informatik in Medizin und Biologie **20**, 3–65.
4. SAS User's Guide: Statistics, Version 5 Edition. SAS Institute Inc. Cary, North Carolina, 403–432.

## F. Eicker

Die Aussicht mag verlockend sein, der Wunsch ist auch nicht neu, Expertenwissen und -können als Nichtexperte auf Knopfdruck maschinell verfügbar zu haben, und in dieser vagen Hoffnung liegt der Grund für die hartnäckige Aktualität des Themas und für die oft nur kurzlebigen Softwareangebote, allen ernststen Warnungen und Vorbehalten von Fachleuten zum Trotz. Auch HAUX warnt zu Recht und eindringlich wie viele vor und neben ihm vor utopischen statistischen Universal-Expertensystemen (XPS), er geht darin sicher nicht zu weit. Man kann nur hoffen, daß die Kritikfähigkeit möglichst vieler Wissenschaftler, von denen die meisten ja gewöhnt sind, die Spreu vom Weizen zu trennen, gegenüber allzu weitgehenden Versprechungen auch auf dem Gebiet der Statistik und ihrer sogenannten XPS geschärft wird. (Am besten sollte man die Bezeichnung »Statistische Expertensysteme«, vom Autor trotz aller Bedenken und trotz guter Alternativen in seinem Aufsatz fast durchgängig benutzt, tunlichst unterdrücken.) Nach dem derzeitigen Gebrauch des Ausdrucks »Statistische XPS« hat man sich darunter wohl enzyklopädische und/oder explanatorische Systeme vorzustellen, die sich auf Modell- und Verfahrensdefinition und -eigenschaften beziehen und somit am ehesten noch den üblichen, theoretisch ausgerichteten statistischen Lehrbüchern ähneln. Solche Systeme würden sich der empirischen Überprüfung und Auffindung mathematischer Modelle in einem bestimmten konkreten Kontext nicht widmen.

Den vielen und gewichtigen Gründen gegen zuvielversprechende statistische XPS, die der Autor anführt oder auf die er in der Literatur verweist, brauchen hier keine weiteren hinzugeführt zu werden; ihr Tenor ist weitgehend der gleiche wie der der wohl- und lange bekannten Warnungen vor mißbräuchlicher oder anfechtbarer Anwendung statistischer (und pseudowissenschaftlicher) Methoden überhaupt.

Einer der möglichen Auswege, zu verlässlicherer KI-Unterstützung zu gelangen, der im Diskussionsartikel auch angedeutet wird, besteht in der Einengung der Softwareentwicklung auf bestimmte Anwendungsfälle. Je enger abgegrenzt ein solcher Fall ist, je bescheidener der Anspruch des Softwaresystems ist, desto vertrauenswürdiger und brauchbarer könnte das Programm werden. Dies ist ein völlig naheliegender Sachverhalt, und er wird auch durch viele positive Beispiele bestätigt: Kleine einfache Reparatursysteme, manche Prozeßrechnerprotokolle fallen hierunter. In solchen Fällen sind die sachlichen Gegebenheiten des Einsatzgebiets hinlänglich bekannt, einiges diesbezügliche Wissen und Schlußfolgerungsregeln mögen sich in den qualitativen Wissensbanken und dem Regelsystem des Programms wiederfinden.

Ein entscheidendes Gegenargument gegen universelle »Statistische XPS«, nämlich das Fehlen des vielfältigen, unverzichtbaren Sachgebietswissens zu all den denkbaren konkreten Anwendungsfällen entfällt in jenen kleinen Spezialsystemen.

Die Akquisition des relevanten Wissens über den empirischen Sachverhalt, der einer statistischen Untersuchung zugrunde liegt, ist keine statistische Aufgabe und müßte bei jedem sachbezogenen XPS durch den jeweiligen Sachgebietsexperten erfolgen. Hier liegt bekanntlich allgemein eine der Hauptschwierigkeiten bei der Realisierung konkreter XPS.

Manchem Statistiker erscheint es angesichts dieser Gegebenheiten ratsam und angemessener, sich darauf zu beschränken, lediglich *Statistikkomponenten* für einen Anwendungsfall in einem bereits erstellten XPS in Betracht zu ziehen und einzubringen; das nötige Hintergrundwissen liegt dann also bereits im System vor. Beispiele sind wiederum Reparatur- sowie gewisse begrenzte Prozeßkontroll- und medizinische XPS. Die Statistikkomponenten könnten sich auf Auswertung, Versuchsplanung, Modellüberprüfung, routinemäßiges oder exploratives Arbeiten, Interpretation der Ergebnisse etc. beziehen. Selbst unter der vorgeschlagenen Eingrenzung des Systems liegt es auf der Hand, daß der Statistikteil noch sehr ausufern könnte.

Einige neuere Entwicklungsarbeiten im Rahmen des »Dortmunder Zentrum für Exper-

tensysteme; Wissensbasierte Systeme in Produktionstechnik und Dienstleistungen«, einem kürzlich erfolgten Zusammenschluß von Hochschullehrern aus sechs Fachbereichen, lassen sich ebenfalls im beschriebenen Sinn als Beiträge von Statistikkomponenten zu bestehenden inhaltsbezogenen XPS der Chemietechnik und zur medizinischen Diagnose verstehen.

Der Fortschritt über den Weg solcher empirischer Spezial-XPS ist zwar langsam, würde im Erfolgsfalle aber zusätzliches Vertrauen wecken und für weitere Schritte werben. Größere sachbezogene wissensbasierte Systeme (z. B. XPS) lassen sich gegebenenfalls mosaikartig aus solchen kleineren, realisierten, fallspezifischen und aufeinanderbezogenen Systemen zusammenbauen. Beispielsweise können so erprobte Diagnosesysteme mit graduell erweitertem Gültigkeitsbereich entwickelt werden. Dahinter stünde die prinzipielle Vorgehensweise, vom kleineren, spezifischen Basisbaustein aufsteigend (induktiv) zu umfassenderen Systemen zu gelangen. (Der umgekehrte Weg, einen großen, abstrakten und leeren Rahmen vorzugeben und diesen absteigend (deduktiv) mit Detailwissen aufzufüllen, wird sicher häufiger vorgeschlagen, ist aber in der Regel nur scheinbar einfacher und im Blick auf die Statistik vielfach nicht empfehlenswert.) Ob es sich im Einzelfall mehr um ein Beratungs- oder mehr um ein Expertensystem handelt, wird sich meist von selbst ergeben und eher zweitrangige Bedeutung haben; ähnliches gilt für den Automatisierungsgrad, die Benutzerklassen, die Zuweisung von Verantwortung bei der Anwendung eines Systems u. v. m. Im übrigen haben Statistikkomponenten in XPS und statistische Methodologie naheliegenderweise viele Aspekte und Kriterien gemein: z. B. den Grad und die Art der Interaktivität, umgekehrt der Automatisierung des Methodeneinsatzes, Benutzerpräferenzen und -fähigkeiten, exploratorisches oder konfirmatorisches Arbeiten, Berücksichtigung empirischen Hintergrundwissens, Transparenz u. v. m. Diese Aspekte werden u. a. in den Abschnitten 4.1 und 5.2 des Artikels von HAUX deutlich angesprochen. Der Autor eines jeden statistikorientierten XPS muß zu jeder dieser seit eh und je diskutierten Frage des Faches Stellung beziehen und sein XPS so oder so ausrichten – es gibt keinen Königsweg zur Vermeidung der Grundfragen, durch KI so wenig wie bei den klassischen Statistikpaketen der Batchzeit. Dies wird sicherlich weitgehend akzeptiert, allerdings wird die Suche nach Wegen, den Mißbrauch zu erschweren, wohl nie enden.

Anschrift des Verfassers: Prof. Dr. F. Eicker, Fachbereich Statistik, Universität Dortmund, Postfach 500500, 4600 Dortmund 50

## K. A. Fröschl

### Zum Begriff

NELDER [8] und CHAMBERS [2] wird im allgemeinen das Verdienst zugeschrieben, zuerst auf die Begegnung von Anwendungsproblemen bei statistischen Programmsystemen mittels »maschineller Intelligenz« hingewiesen zu haben; allerdings findet sich schon ein wenig früher bei FINNEY [5] (p. 17) folgende Feststellung: »We need to ensure that *protection* is built into the more popular programs, perhaps by incorporating into each an *internal monitor* that can identify some of the more common flaws in the *quality of data* and in the *appropriateness of standard models*.« Darüber hinaus hat FINNEY eine »... more ambitious idea« betreffend die Entwicklung statistischer Programme mit »... built-in statistical *advice to guide* the user in avoiding the more serious mistakes of methodology«. (Kursivstellungen nicht im Original.) Damit sind die im Zentrum der aktuellen Diskussion stehenden Motive der Gestaltung statistischer Programmsysteme unter dem Einfluß von Ideen und Konzepten

der sogenannten *Künstlichen Intelligenz*<sup>1</sup> – freilich noch ohne Rekurs auf inzwischen sattsam bekannte magische Modewörter – unmittelbar angesprochen.

Trotz dieses klar artikulierten Problembewußtseins ist in den vergangenen eineinhalb Dezennien die Antwortfindung jedoch eher zögernd vorangeschritten. Dank der sprießenden Popularität von *Expertensystemen* firmieren praktisch alle einschlägigen Software-Entwicklungen zwar unter diesem »trendy«-Etikett, aber es ist noch nicht einmal klar, was denn eigentlich unter »statistischen Expertensystemen«<sup>2</sup> genauer verstanden werden soll<sup>3</sup>. Im folgenden werden zwei – einander nicht unbedingt ausschließende, aber doch in der Blickrichtung entgegengesetzte – Varianten der Begriffsauslegung diskutiert, und zwar (i) SES als Simulation/Modell des angewandten Statistikers bzw. (ii) SES als auf die Statistik angewandte Expertensystem-Technologie.

*ad (i)*: Zunächst erfordert jede Simulation die Spezifikation des Abstraktionsgrades (d. h. der Detailtreue) des Modells relativ zur Realität. Die Entwicklung von SES vollzieht sich solchermaßen vor einem Spektrum mit den Polen »Expertenmodellierungs-System« und »Expertenleistungs-System« (vgl. [3]). Dieses plakative Bild wird aber durch zwei Aspekte unterlaufen; einerseits ist gut vorstellbar, daß eine zufriedenstellende Expertenleistung solcher Systeme ausschließlich auf dem Wege einer hinreichend mit Detailstruktur ausgestatteten Modellierung erreicht werden kann (d. h. Black Box-Modelle sind zu stumpf), andererseits reduziert die KI diese Modellierung von vornherein auf eine *linguistische* Ebene (d. h. die Simulation von Informationsverarbeitungsprozessen in der Sprache der Logik). Wenn man bereit ist zu unterstellen, daß in der Statistik (bzw. allgemeiner: Empirie) auch andere als *verbalisierbare* – und über diese hinausgehende – kognitive Funktionen eine wesentliche Rolle spielen, dann bleibt das Ziel einer Expertenmodellierung mit den Mitteln der KI notwendig relativ.<sup>4</sup> Die (formale) Sprache ist indes der einzige »sensorische Eingangskanal« der Maschine; insofern statistische Gedankeninhalte sprachlich formulierbar sind, hat die Entwicklung künstlich intelligenter Statistiksysteeme aber durchaus prinzipielle Chancen auf Erfolg.<sup>5</sup>

*ad (ii)*: Die populärere und zumeist stillschweigend präferierte Interpretation betrachtet SES vor dem Spektrum datenanalytischer *Dienstleistungen* unterschiedlichen Automationsgrades, an dessen Enden vollautomatische Datenexploration (System ist initiativ) und die – passive – Werkzeugkiste CUS = Computer Unterstützte Statistik einander gegenüberliegen. Gekennzeichnet ist dieses Spektrum durch ein Spezifikationsgefälle der systeminternen Repräsentation statistischer Probleme: die automatische Analyse<sup>6</sup> geht von eher vage formulierten, im allgemeinen keine eindeutige Lösung determinierenden Problemen aus; Werkzeugsysteme andererseits »internalisieren« die zu lösenden Probleme gar nicht. Im mittleren Bereich des Spektrums hingegen beginnt die Systemaktivität bei klar definierten Problemen und entspricht damit dem schon klassischen KI-Paradigma des *problem solving*: Expertensysteme sind als »mechanische Konsulenten« genau dafür konzipiert, in ex ante abgegrenzten Problembereichen Antworten zu präzise gestellten Fragen zu suchen. Nach dieser Begriffsauslegung ist die Entwicklung von SES dann sinnvoll und aussichtsreich,

<sup>1</sup> Kurz: KI

<sup>2</sup> Künftig kurz mit SES bezeichnet.

<sup>3</sup> Auch der von HAUX (Absatz 3.1) unternommene Versuch einer Begriffsbestimmung dürfte nur eine geringe Halbwertszeit aufweisen: die Formel »spezifische Expertentätigkeiten« ist selbst recht unspezifisch; außerdem werden diese Tätigkeiten naturgemäß (und dialektisch) auch von den verwendeten Werkzeugen – wie Expertensystemen? – geprägt. Konkreter, wenn nicht konstitutiv für derartige Systeme ist der Aspekt (die Absicht) einer formalen Modellierung statistischer *Entscheidungslogik*. In einem umfassenden Sinn ist das Kriterium intelligenter Statistiksysteeme eine »innere Repräsentation« der statistischen *Außenwelt*.

<sup>4</sup> Die Alternative neuronaler, kybernetischer Modelle ist zur Zeit wohl – trotz ihres Revivals – noch immer viel zu wenig attraktiv für die Simulation *abstrakter* Informationsverarbeitungsprozesse; vgl. z. B. CHANGEUX [4], Kapitel 5. Siehe etwa auch HAMEROFF [7].

<sup>5</sup> Dies ungeachtet formallogischer Gegenargumente, wie sie etwa STREITBERG vorträgt; vgl. [10] und die anschließende Diskussion. Die dort angesprochenen Probleme sind eher Indizien für eine *inadäquate* Modellierung.

<sup>6</sup> Selbstverständlich wird dieser provokanten Formulierung mit Skepsis begegnet. In letzter Konsequenz scheint eine solche Konzeption auch unproduktiv: die empirische Forschung folgt gesellschaftlichen, nur vordergründig wissenschaftsimmanenten Bewertungskriterien, die von den – bislang jedenfalls – *asozialen* Maschinen nicht nachvollzogen werden können.

wenn in (Teilbereichen) der Statistik Problemfelder so definiert werden (können), daß sie der Funktionalität von Expertensystemen entgegenkommen. Wie so oft im Zusammenhang mit der Digitalisierung läßt sich eine (mehr oder weniger bewußte?) Revision statistischer Auffassungen über die Analyse empirischer Daten im Zuge dieser »Anpassung« allerdings nicht ausschließen.

### Argumente pro

Der Allgemeinheit willen sei im weiteren mit SES die Gesamtheit aller unter dem Einfluß und Eindruck der KI entwickelten statistischen Programmsysteme gemeint. Diese Begriffs-*entleerung* rechtfertigt sich durch die potentielle Fruchtbarkeit einer möglichst unvoreingenommenen Betrachtung, die etwa die folgenden Nutzenmotive umschließt:<sup>7</sup>

**Ethisches Motiv:** Durch die Integration *statistischer Intelligenz* in die vornehmlich numerisch orientierten Programmsysteme könnte der Kritik an der – durch die in dieser Hinsicht fahrlässig gestalteten »Softwarepakete« zweifellos begünstigen – (unbewußten) Fehlverwendung statistischer Verfahren bzw. einer generell mangelhaften statistischen Methodik entgegengetreten und solcherart das allgemeine Niveau statistischer Argumentation *im ureigensten Interesse der Statistik* angehoben werden.

**Praktisches Motiv:** Geeignet genutzt, enthält der Digitalrechner das Potential eines statistischen »Allround-Mediums«, das alle erdenklichen Hilfsmittel zur Durchführung statistischer Datenauswertungen in sich vereinigt. Denkbare Erweiterungen der konventionellen Funktionen statistischer Programmsysteme stellen etwa (vgl. [6]) Komponenten zur computerisierten Literatur- und Programmberatung, zur systemgestützten Organisation gesamter Datenauswertungen und zur kooperativen, fallbezogenen Methodenberatung dar, die vor allem durch ihre *Vernetzung* eine neue Qualität in die statistische Arbeitsroutine bringen könnten.

**Methodologisches Motiv:** Die Entwicklung intelligenter Statistiksysteme forciert – durch den Zwang zur formalen Präzisierung – die *statistische Eigenanalyse*. In diesem Sinn kehrt sich der konventionelle Zusammenhang von SES und Statistik um: SES werden zu einem *experimentellen* Werkzeug der strukturalistischen Analyse angewandter Statistik, zu einem Hilfsmittel innerstatistischer Reflexion gewissermaßen, aus der nicht nur fundiertere Prinzipien zur nachfolgenden Gestaltung von SES, sondern gleichzeitig Anstöße zur Ausfüllung von Theorielücken durch die mathematische Statistik resultieren. Das letztendliche Ziel einer solchen Eigenanalyse ist die Formulierung einer kohärenten *rationalen Theorie* statistischer Argumentation.

**Epistemologisches Motiv:** Vom alltagspraktischen Standpunkt vielleicht weniger vordringlich, aber in ihrer Konsequenz für die Rolle der Statistik im wissenschaftlichen Erkenntnisprozeß letztlich von eminenter Bedeutung ist die Frage des Einflusses intelligenter Statistiksysteme auf Dateninterpretation, Modell- und Theorienbildung. Nimmt man das philosophische Programm der KI beim Wort (und setzt voraus, daß die KI ihre Ziele im wesentlichen realisieren kann), dann sind intelligente Statistiksysteme auf lange Sicht ein wertvolles *konkretes* Analysewerkzeug zur experimentellen Untersuchung von Hypothesenbildung und Erkenntnisgewinnung und damit eine qualitativ völlig neuartige Alternative wissenschaftstheoretischer Argumentation (die selbstredend auf die KI zurückwirkt).

**Technologisches Motiv:** Die Zielsetzung der Entwicklung intelligenter Statistiksysteme bedingt nicht nur eine Prüfung der Tragfähigkeit allgemeinerer, in der KI gewachsener Realisierungsmodelle, sondern bringt selbst eine Reihe spezifischer Probleme in die Thematik *wissensbasierter* Programmsysteme. Einer synergistischen Auseinandersetzung mit den beim Entwurf von SES auftretenden Problemen könnten durchaus beiden Disziplinen, Statistik und KI, zugute kommende Fortschritte und Einsichten entspringen.

<sup>7</sup> Eine ausführlichere Diskussion dieser Motive enthält meine in Vorbereitung befindliche Dissertation.

### Argumente kontra

Betrachtet man den »Stand der Kunst« in der Entwicklung (in der Auffassung der KI) intelligenter Systeme bzw. Expertensysteme, so steht man bereits vor dem gegenwärtig augenscheinlichsten Argument *wider* die Möglichkeit der Realisierung gediegener SES, denn im Sinne etwa der Terminologie von BUCHANAN [1] entsprechen sinnvolle SES viel eher den »Level 2«-Systemen mit einer Reihe von Merkmalen, wie die Einbeziehung von kausalen und strategischen Modellen, zeitlichen und räumlichen Konzepten, Begriffsabstraktionen, Analogien, Konfliktresolution in mehrdeutigen Entscheidungssituationen, und nicht zuletzt: Erfahrungslernen, die den bisher machbaren »Level 1«-Systemen noch weitestgehend fehlen. Vor allem der letzte Aspekt – Erfahrungslernen – stellt ein größeres Handicap heutiger Expertensysteme dar; als starre, nichtkybernetische Automaten mögen sie einmal fixiertes Expertenwissen besser einsehbar, beurteilbar und *verteilbar* machen, aber dem ökonomischen Nutzen des breiten Systemeinsatzes (»Expertenduplikation«) stehen die Gefahren einer mangelnden Flexibilität bzw. *Kreativität* und einer tendenziell *verlangsamten* Umsetzung innovativen Wissens gegenüber.

Konkret auf die angewandte Statistik bezogen, muß aus dieser Situation wohl bis auf weiteres der Schluß gezogen werden, daß SES am ehesten für standardisierte, *stereotype* Schablonenaktivitäten formuliert werden können. Solche Bereiche finden sich in der Statistik zweifellos, aber es bleibt die Frage, ob die Systeme dann auch den ursprünglich hoch gesteckten Zielen (siehe oben) gerecht werden. NELDER [9] gibt zudem zu bedenken, daß – in deutlichem Gegensatz zu den konkreten Aufgaben anderer Expertensysteme – SES auf abstrakte Klassen von Problemen ausgelegt werden müssen, weil ein Statistiker (von Ausnahmen abgesehen) nicht Routineprobleme löst, sondern zuerst einmal die Anwendbarkeit von Routinestrategien auf höchst individuelle Probleme *überprüft*.<sup>8</sup> Der Aufgabenbereich von SES ist damit ex definitione erheblich umfangreicher als bei Expertensystemen sonst üblich.

Schließlich ist die typische Anwendungssituation von Expertensystemen – suche Lösung zu festem Problem mit variablen Parametern – im Bereich der Statistik an sich hochgradig kontraindiziert. VELLEMAN [11] etwa meint überhaupt, das Konzept von SES im Sinne mechanischer Konsulenten könnte schlicht falsch sein und ersetzt es durch das Modell eines »Forschungsgehilfen«. Damit scheint jedoch der Übergang von im eigentlichen Sinn *problemlösenden* SES zu *entscheidungsunterstützenden*, in einem tieferen Sinn kooperativen Systemen, in denen die Rollen der besseren Kompetenz auf Mensch und Maschine komplementär verteilt sind, bereits vollzogen. Wenn VELLEMAN den Kern der Sache getroffen hat (und vieles spricht dafür), dann sollte man in der Tat von SES (in der engeren Deutung des Begriffs) nicht sehr viel Nutzen erwarten, weder für den unroutinierten und ergo anleitungsbedürftigen Verwender statistischer Software, noch gar für einen – einigermaßen – kompetenten Statistikanwender. Es bliebe diesfalls lediglich die Frage: wie lassen sich »SES« wirksam verhindern? Denn: daß *solche* Systeme auch entgegen allen statistischen Vorbehalten realisiert werden, ergibt sich schon aus der Logik des Bedarfs, i.e. eines praktisch nicht zu vernachlässigenden Nachfragemarktes.

### Zur Situation

Konventionelle statistische Software-Systeme weisen unbestrittenermaßen mehr oder weniger gravierende Defizite auf, sei es nun hinsichtlich der Form der Kommunikation zwischen Mensch und Maschine oder sei es hinsichtlich der Flexibilität der Abbildung inhaltlicher statistischer Probleme auf die vordefinierten Sprachstrukturen der Systeme. Die KI und deren Apologeten, die oft genug selbstbewußt die kognitiven Fähigkeiten ihrer maschinell-

<sup>8</sup> Das entspricht, wenn man so will, einer *heuristischen* Minimierung des Fehlers »dritter Art«, wie ihn auch HAUX erwähnt.

len Modelle anpreisen, haben eine Weiterentwicklung der statistischen Software in dieser Richtung angeregt. Anstelle die Konzepte der KI in einer ersten Euphorie unreflektiert als General-Heilmittel allfälliger Mängel in der computationalen Statistik zu proklamieren, empfiehlt sich zunächst einmal eine gründliche Analyse von Mängelursachen; in der weiteren Folge erst läßt sich beurteilen, in welcher Weise die KI den inneren Gegebenheiten der statistischen Datenverarbeitung am nützlichsten Rechnung trägt. Gerade im Zusammenhang mit SES scheint es wichtig gelegentlich darauf hinzuweisen, daß im Gegensatz zur meist in den Vordergrund gestellten – technischen – Frage der statistischen Entscheidungslogik (Regelsystem aufstellen usw.) die viel wichtigeren Fragen nach dem Ursprung<sup>9</sup> und der Verwaltung der für sinnvolle Entscheidungen erforderlichen Information viel zu wenig Beachtung finden.

Generell ist anzumerken, daß die Möglichkeit einer Verwendung von SES – ungeachtet der Begriffsauslegung im einzelnen – durch *Nichtexperten* eher zurückhaltend beurteilt werden muß; die Verwendung (auch noch so) »intelligenter« Systeme durch Endbenutzer, deren eigene Fachkundigkeit de facto nicht zu einer Einschätzung der von solchen Systemen produzierten Lösungen und Vorschläge ausreicht, wirft besonders vehement die Frage der Verantwortung falscher Entscheidungen auf. Nach dem gegenwärtigen Stand der Entwicklung sind SES (im weitesten Sinn) eher als Werkzeug selbst statistisch fachkundiger Verwender denkbar. In einem solchen Szenario computergestützter Statistik gibt der Systemverwender seine Kompetenz und Verantwortung keineswegs an die Maschine ab, aber diese kann ihm als »intelligente« Arbeitshilfe wirkungsvoll helfen, seine Kompetenz besser zu entfalten und seine Verantwortung stärker abzusichern. In dieser Auffassung liegt längerfristig wohl das größte Nutzenpotential statistischer Expertensysteme.

## Literatur

1. BUCHANAN, B. G. (1982): New research on expert systems. In: HAYES, J. E., D. MICHIE, Y.-H. PAO (Hrsg.) *Machine Intelligence 10*, Ellis Horwood, Chichester, pp. 269–309.
2. CHAMBERS, J. M. (1981): Some thoughts on expert software. In: EDDY, W. F. (Hrsg.) *Computer Science and Statistics: Proc. 13<sup>th</sup> Symp. on the Interface*, Springer, New York, pp. 36–40.
3. CHAMBERS, J. M., W. A. GALE, D. PREGIBON (1988): On the existence of expert systems (Kommentar zu [10]). *Stat. Softw. Newsletter* 14 (2), pp. 63–66.
4. CHANGEUX, J.-P. (1983): *L'homme neuronal*. Librairie Arthème Fayard, Paris.
5. FINNEY, D. J. (1974): Problems, data, and inference. *J. R. Stat. Soc.* A137 (Part 1), pp. 1–23.
6. HAHN, G. J. (1985): More intelligent statistical software and statistical expert systems: future directions, *American Statistician* 39 (1), pp. 1–8.
7. HAMEROFF, S. R. (1987): *Ultimate Computing*, North Holland, Amsterdam.
8. NELDER, J. A. (1977): Intelligent Programs, the Next Stage in Statistical Computing. In: BARRA, J. R. et al. (Hrsg.) *Recent Developments in Statistics*, North Holland, Amsterdam, pp. 79–86.
9. NELDER, J. A. (1988): The Rôle of Expert Systems in Statistics. In: FAULBAUM, F., H.-M. UEHLINGER (Hrsg.) *Fortschritte der Statistik-Software I*. G. Fischer, Stuttgart, pp. 175–182.
10. STREITBERG, B. (1988): On the nonexistence of expert systems – critical remarks on artificial intelligence in statistics. *Stat. Softw. Newsletter* 14 (2), pp. 55–62.
11. VELLEMAN, P. F. (1985): Kommentar zu [6]. *American Statistician* 39 (1), pp. 10–11.

Anschrift des Verfassers: Karl Anton Fröschl, Institut für Statistik und Informatik, Universität Wien, Universitätsstraße 5/9, A-1010 Wien

<sup>9</sup> Wie sind Fragen an den Systemverwender am besten zu formulieren bzw. zu gestalten? In welcher Reihenfolge sollen Fragen gestellt werden? Wie sind suggestive Formulierungen vermeidbar? Welche Struktur hat ein »statistischer Dialog«? Wie veranlaßt man den Systemverwender, (möglichst nur) substantielle Information bereitzustellen? Wie kann für den Systemverwender die Möglichkeit sichergestellt werden, die ihm wesentlich scheinenden Sachverhalte auch wirklich dem System mitzuteilen? Usw. usf.

## K.-H. Jöckel, R. Ostermann

Mit dem in der Arbeit von R. HAUX gegebenen Überblick über den derzeitigen Stand der Entwicklung statistischer Expertensysteme ist der Versuch gelungen, dieses unter Statistikern/Biometrikern kontrovers diskutierte Thema wieder einer sachlichen wissenschaftlichen Diskussion zu öffnen. Die präzise Darstellung der Grundbegriffe und der Funktionalität solcher Systeme erlaubt den Verfassern dieses Kommentars, sich auf einige Teilaspekte zu beschränken.

Stärkere Beachtung verdient u. E. in der Diskussion um statistische Expertensysteme der Aspekt der Mensch-Maschine-Interaktion, und zwar der des Benutzers mit dem System. Auch wenn HAUX die Benutzung von statistischen Expertensystemen durch statistische Laien als problematisch ansieht, muß davon ausgegangen werden, daß solche Systeme, ähnlich wie bereits heute die statistischen Auswertungssysteme, auch diesem Nutzerkreis offenstehen werden. In welcher Weise dann Fehler der »dritten« Art vermieden werden können, hängt nicht zuletzt davon ab, wie ein statistisches Expertensystem geeignet mit dem Nutzer interagiert. Daher sollten sich empirisch arbeitende Wissenschaftler nicht mit theoretischen Überlegungen zur Adäquanz des einen oder anderen Systemverhaltens zufriedengeben, sondern sich vielmehr bemühen, empirische Daten, z. B. mit Hilfe psychologischer Experimente, zu erhalten. Auf diesen Aspekt ist bereits von MOLENAAR, BROERSMA (1984) für statistische Auswertungssysteme hingewiesen worden.

Zu Recht wird von HAUX der Aspekt der Reproduzierbarkeit wissenschaftlicher Problemlösungen hervorgehoben. Neben der theoretischen Herleitbarkeit wissenschaftlicher Forschungsstrategien spielen aber (dem Statistiker häufig unbewußt) auch Erfahrungswerte eine wesentliche Rolle. Statistisches Wissen ist eben nicht nur Wissen über die mathematisch-statistischen Eigenschaften und abgeleitete Adäquanz bestimmter Strategien (z. B. Doppelblindversuch), sondern auch empirisches Wissen über den Erfolg ganz bestimmter Vorgehensweisen. Besonders deutlich wird dies im Bereich der Prognosen, wo neben theoretischen Überlegungen auch empirische Ergebnisse zum Abschneiden ganz bestimmter Prognoseverfahren vorliegen, vgl. NEWBOLD, GRANGER (1974). Ein statistisches Expertensystem zur Prognose beispielsweise volkswirtschaftlicher Zielparameter könnte durchaus adaptiv konstruiert in dem Sinne sein, daß das Abschneiden bestimmter Prognoseverfahren in der Vergangenheit mit für die Auswahl des Verfahrens zur Abgabe einer neuen Prognose herangezogen wird. Da hinlänglich bekannt ist, daß simple Verfahren wie das einfache »play the winner« eine schlechte Strategie darstellen, ergeben sich hier schon bei der Konstruktion eines Expertensystems wesentliche Arbeitsbereiche für den methodisch orientierten Statistiker. Ein nichtautoritäres System, das solche Inferenzmechanismen inkooperiert hat, könnte bei der Auswahl eines konkreten Prognosesystems durchaus mit Erfahrungen aus der Vergangenheit argumentieren, dieses im Unterschied zu einem menschlichen Experten aber auch quantitativ belegen.

Ein weiterer Beitrag zur Reproduzierbarkeit statistischer Auswertungsansätze könnte weiterhin darin bestehen, daß ein Expertensystem seinen (statistischen) Nutzer auf mögliche Inkonsequenzen bei der Auswahl statistischer Verfahren hinweist. Zu denken ist dabei beispielsweise an Situationen, in denen der Statistiker üblicherweise nichtparametrische Verfahren eingesetzt hat, in einer konkreten neuen Anwendung jedoch den t-Test verwendet. Als einen in diesem Zusammenhang wichtigen Begriff wollen wir den Terminus »Lab-Assistent« erwähnen (HUBER (1986)). Dieser Lab-Assistent soll sogenannte Script-Files (also Protokolle einer Datenanalyse-Sitzung am Computer) auswerten. Die Auswertung dient der Aufbereitung des Materials zur Reporterstellung, der Überprüfung der Korrektheit der durchgeführten Analysen, zur Wiederholung der Analyse mit modifizierten Daten oder Methoden. Interessante Ausführungen zur Analyse dieser Script-Files findet man u. a. in VACH, HUBER (1988). Insbesondere bei der Planung klinischer Studien wären hier nicht zu vernachlässigende Standardisierungseffekte zu erzielen, vgl. JÖCKEL (1987).

Der Aspekt des Zusammenhangs zwischen statistischen Expertensystemen und Datenbanksystemen verdient über den von HAUX angesprochenen »robot-statistician« hinaus noch mehr Beachtung. Es entspricht dem üblichen Verständnis innerhalb der Statistik, daß man bei einer Analyse zwischen Primärdaten (d. h. für den Studienzweck gesammelte) und Sekundärdaten (d. h. für einen anderen Studienzweck gesammelte) unterscheidet. Die möglichen Konsequenzen einer falschen Verwendung von Sekundärdaten werden von HAUX deutlich aufgezeigt. Dennoch haben Daten für den empirisch arbeitenden Statistiker neben der Beantwortung neuer Fragestellungen auch den Zweck der Überprüfung der Adäquanz statistischer Verfahren an bereits ausgewerteten Datenbeständen. Ein berühmtes Beispiel stellen die in der Literatur wieder und wieder analysierten Daten von FISHER (1936) zu bestimmten Charakteristika der Blattgröße bei Iris dar. Solche historische Daten oder aber auch simulierte Daten zur Beurteilung der Leistungsfähigkeit einer bestimmten Forschungsstrategie heranzuziehen, ist ein wohletabliertes Instrument der Statistik, darüber hinaus liefern solche historische Datenbestände häufig wesentliche Planungsgrößen (z. B. Standardabweichung) für die Versuchsplanung. Insofern bleibt zu überlegen, inwieweit zukünftige Expertensysteme Inferenzmechanismen benutzen können, die Erfahrungen aus der Analyse bekannter Datenbestände einbeziehen, vgl. JÖCKEL (1986). Die Ergebnisse von Simulationsstudien, vgl. z. B. zur Problematik der Stichprobenumfangbestimmung JÖCKEL (1979), aber auch Techniken wie das Bootstrapping JÖCKEL (1987) ließen sich hierunter subsumieren.

Ein letzter Kommentar bezieht sich auf die von HAUX beschriebene Methodenbankhülle SAMBA/PC als wissensbasierte Schnittstelle, mit deren Qualität er bei den dort gegebenen Interpretationshilfen nicht zufrieden ist. So ist den Autoren von SAMBA/PC die Schwierigkeit einer sauber formulierten Interpretationshilfe vollauf bewußt (BODENDORF (1988)), jedoch muß ein jeder Autor, der einmal versucht hat, Gleichartiges zu formulieren, die Erfahrung machen, wie schwierig es ist, derartige Formulierungen auszuarbeiten. Dennoch stimmen wir bzgl. der schwachen Qualität mit HAUX voll und ganz überein. Was jedoch noch sehr viel interessanter erscheint, ist die Tatsache, daß sich ein derartiges Software-Produkt am Markt nicht durchsetzt, d. h. es sich »für eine professionelle Verwertung als ungeeignet herausgestellt hat« (RAUH (1988)). Auch im System SAMBA/PC sollen gewisse Entscheidungen (Normalverteilung: ja oder nein) nach Möglichkeit vom Benutzer getroffen werden. Damit wird auf jeden Fall das Problem der multiplen Tests umgangen.

## Literatur

- BODENDORF, F. (1988): Persönliche Kommunikation.  
 FISHER, R. A. (1936): The Use of Multiple Measurements in Taxonomic Problems, *Ann. Eugen.* **8**, 376–386.  
 HUBER, P. J. (1986): Environments for Supporting Statistical Strategy in: Gale, W. A. (Hrsg.), *Artificial Intelligence*, Addison-Wesley.  
 JÖCKEL, K.-H. (1979): Sample Size Determination by Simulation, in: *EDV in Medizin und Biologie*, **10**, 58–61.  
 JÖCKEL, K.-H. (1986): Statistical Expert Systems and the Statistical Consultant—Considerations about the Planning Stage of Clinical Studies, in: HAUX, R. (Hrsg.), *Expert Systems in Statistics, Selected Papers from a Workshop*, 27–43.  
 JÖCKEL, K.-H. (1987): Monte Carlo Techniques and Hypothesis Testing. Invited Paper at the First International Conference on Statistical Computing (ICOSCO-I) erscheint.  
 MOLENAAR, I. W., BROERSMA, H. (1984): Interpretation of Statistical Software Output: Some Behavioural Studies, in: *COMPSTAT 1984*, Physika-Verlag Wien, 341–346.  
 NEWBOLD, P., GRANGER, C. W. J. (1974): Experiences with Forecasting Univariate Time Series and the Combination of Frequencies, *J. Royal Stat. S. A.* **137**, 131–165.  
 RAUH, N. (1988), Persönliche Kommunikation,

VACH, W., HUBER, P. J. (1988), Analysis of ISP-Script-Files as an Example for the Retrospective Analysis of Machine-Written Records, Forschungsbericht 3/88, FB Statistik, Universität Dortmund.

Anschrift der Autoren: PD Dr. rer. nat. K.-H. Jöckel, Bremer Institut für Präventionsforschung und Sozialmedizin (BIPS), St.-Jürgen-Str. 1, 2800 Bremen 1, Dr. rer. nat. R. Ostermann, Hochschulrechenzentrum, Universität - Gesamthochschule - Siegen, Hölderlinstr. 3, 5900 Siegen

## H.-J. Lenz

Auf dem überlappenden Gebiet von Statistik und Informatik gibt es wohl neben der Graphik keinen Bereich, der im letzten Jahrzehnt so viel Attraktivität im Wissenschaftsbereich erworben hat wie der Bereich der statistischen Expertensysteme. Insofern ist es höchste Zeit für eine bewertende Bestandsaufnahme.

Es ist das Verdienst von R. HAUX, der über langjährige Erfahrungen in der Entwicklung und Beurteilung von Expertensystemen verfügt, eine erste Sichtung des Stands der Entwicklung von Expertensystemen durchgeführt und Perspektiven möglicher Entwicklungen aufgezeigt zu haben.

Bekanntlich ist es stets einfacher, auf der Basis eines »Grundlagenpapiers« zu argumentieren, als eine Vorlage zu erstellen. In diesem Sinne sollen die folgenden Ausführungen stets im Lichte konstruktiver Ergänzung der HAUXschen Thematisierung gesehen werden. Um die schnelle Vergleichbarkeit der Erörterungen zu sichern, folgt dieser Kommentar der sachlichen Gliederung der Arbeit von HAUX.

### 1. Begriffsbestimmung

Daß der Forschungsgegenstand »Statistische Expertensysteme« eine junge Disziplin sein muß, erkennt man in der einschlägigen Literatur sowie in der Arbeit von HAUX an der Variationsbreite der Definitionen, was denn nun ein Expertensystem überhaupt sein soll.

Die Erfahrung lehrt, daß inhaltliche Umschreibungen nicht weiterführen. Man denke nur an den Begriff »Wahrscheinlichkeit« im Sinne von Wettquotienten. Erst die formale Begriffsbestimmung als Abbildung  $P: \mathcal{A} \rightarrow \mathbb{R}$ , die den Kolmogoroffschen Axiomen genügt, wobei  $(\Omega, \mathcal{A})$  der Maßraum eines Zufallsexperiments ist, fand allgemeine Zustimmung. Dies läßt sich an den Begriffen »Algorithmus«, »Datenbanksystem« usw. ebenfalls instruktiv illustrieren.

So verlagert HAUX Begriffsbestimmung mittels »eine[r] Maschine (. . .) [die] Funktionen, die für einen Experten spezifisch sind, übernehmen kann« (S. 4). Was sind denn dies für Funktionen, die für einen Experten typisch sind?

Wenn man polemisch antworten würde, könnte man zuerst an »hoher Stundensatz« oder »hohe Honorarforderung« denken. Auch menschliche Experten machen Fehler, sog. »Kunstfehler«. Gehören die auch zu den genannten Funktionen? Aus dem Kontext kann man schließen, daß es dem Autor eher auf die Tätigkeiten selbst ankommt, die Experten ausführen. Überhaupt nicht trivial ist aber die Antwort, worin denn nun diese Tätigkeiten genau bestehen. Der geneigte (menschliche) Leser wird diese Frage für trivial halten, da er sie in menschlich genialer Weise kontextabhängig ergänzt, wobei er aber mit Sicherheit zu verschiedenen Begriffsextensionen und -intensionen kommen dürfte, oder sie als überzogen in der Sicht verdammt.

Da die Analogie zwischen menschlichen Experten und maschinellen Experten von konstitutiver Bedeutung zu sein scheint, ähnlich der Situation in der Robotik, sollen hier einige relevante Aspekte genannt werden, die in jedem Fall die Architektur eines statisti-

schen Expertensystems (kurz: StXPS) beeinflussen dürften, vgl. NAEVE, STEINECKER (1977):

- (1) Statistische Experten verfügen über *Lehrbuch- und Methodenwissen*. Dies Wissen umfaßt präzise Definitionen, Prämissen und Sätze (Theoreme) sowie Heuristiken und Metawissen. Es wird noch ergänzt durch algorithmisches, datenmodellmäßiges, numerisches und Hardware-Wissen. Es sollte vollständig (?), widerspruchsfrei und state-of-the-art sein.
- (2) Statistische Experten haben die *Fähigkeit zum Beraten*, d. h. mit Anwendern im Dialog zu einem substanzwissenschaftlichen Problem (mindestens) ein Modell und (mindestens) eine Methode zur Problemlösung zu finden. Das »eigentliche statistische Problem« zu finden, setzt neben Methodenwissen *Anwendungs- oder Erfahrungswissen* des Statistikers sowie eine gewisse Sprachregelung zwischen Statistiker und Substanzwissenschaftler voraus. Hinzu kommt die Fähigkeit, sich Wissens- oder Datenquellen zu erschließen. Schließlich sind die Lösungsidee und die Schlußfolgerung dem Klienten zu vermitteln.
- (3) Statistische Experten müssen didaktisch geschickt sein, um die auf Versuch und Irrtum basierende Konsultation reibungsfrei abzuwickeln. Am Ende einer Konsultation steht i. a. »tiefere Einsicht« in das substanzwissenschaftliche Problem beim Klienten und breiteres Erfahrungswissen beim Statistiker.

Da es unzweckmäßig erscheint, eine Definition des Begriffs »Statistische Expertensysteme« mit Attributen des menschlichen Experten in der Statistik zu überfrachten, scheint eine formale Definition zweckmäßiger. In Anlehnung an RICHTER (1986) wird als Expertensystem (XPS) definiert:

Das Quintupel (I, T, O, DB, KB) heißt Expertensystem (XPS), wobei

I = Menge aller Inputs (Klientenanfragen)

T = Menge aller Transformationen (Inferenz, Datenoperationen relationaler Art und -funktionen)

O = Menge aller Outputs (Systemantworten)

DB = (relationales) Datenbanksystem mit Menge von Tabellen

KB = Wissensbank mit Menge von Fakten, Regeln/Rahmen/prädikatenlogischer Aussagen.

Das Expertensystem wird zum Statistik-Expertensystem, falls DB bzw. KB eine entsprechende Semantik aufweisen und zur Modellverwaltung und zum Methodenzugriff das XPS um ein Statistikprogrammpaket erweitert wird. Offensichtlich spielen die Komponenten

- Daten
- Modell
- Methode
- Modellierungsunterstützung
- Ergebnisdarstellung

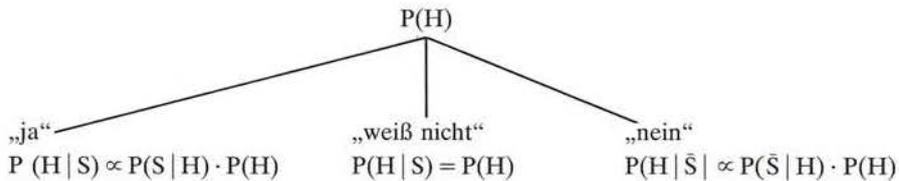
die Hauptrolle.

## 2. Funktionsweise

Expertensysteme sind auf bestimmte Anwendungsgebiete (domains) zugeschnitten, für die sie Kompetenz besitzen. Von dem Strom der Beratungsfälle, der auf ein Expertensystem zukommt, ist zu verlangen, daß er »homogen« ist; dieser Begriff bedarf sicherlich der Präzisierung. Die Situation ist vergleichbar beispielsweise einem Expertensystem »Motortester«, das für Ottomotoren in Aluminiumbauweise ausgelegt ist. Man stelle sich vor, in einem Strom zu testender Motoren mit üblicherweise 1600 cm<sup>3</sup> bis 4500 cm<sup>3</sup> tauche ein Motor mit 50 cm<sup>3</sup> auf. Wie wird das XPS reagieren? Es scheint, daß bei der Beratung auf dem Gebiet des Einsatzes statistischer Methoden a priori nicht klar ist, ob das Problem im

Kompetenzrahmen des betreffenden Expertensystems liegt oder nicht. Die Kompetenzprüfung erfolgt daher im allgemeinen außerhalb des XPSs.

Ein weiteres Problem beruht auf nichtdeterministischen Inferenzmechanismen, die selbstanpassend sind, wie es mit dem BAYES-Konzept realisiert werden kann. Will man damit ein Diagnosesystem beschreiben, so könnte man wie folgt vorgehen:



H: Hypothese

S: Symptom

Ein Kinderarzt könnte bei einem Kind an einem bestimmten Tag eine Grippe diagnostizieren. Hat das nächste behandelte Kind ebenfalls Grippe, so wird er wohl bei jedem neuen Patienten Grippe als erste zu prüfende Krankheitsursache vermuten, da eine Epidemie vorliegen könnte. Bei negativen Tests würde er seine Meinung (Präjudiz) korrigieren.

Man sieht jedoch, daß  $\lim_{k \rightarrow \infty} P_k(H|S) = 1$  für entsprechende Beobachtungsfolgen  $k = 1, 2, \dots$ . Auch bei kleinem  $k$  kann die Beobachtung von  $\bar{S}$  das Expertensystem nicht mehr von der Regel  $S \xrightarrow{P(H|S)=1} H$  wegführen. Der absorbierende Zustand verhindert jedes weitere Lernen. Dies ist im offensichtlichen Gegensatz zum Verhalten eines Diagnostikers, der irgendwann das Ende einer Epidemie konstatiert und entsprechend handelt und – technisch gesprochen – sein  $P(H|S)$  „zurücksetzt“ auf  $0 < P(H) < 1$ .

### 3. Repräsentationsformen

Prädikatenlogik, Produktionsregeln, Rahmen (Frames) und semantische Netze sind die am häufigsten genannten Paradigmen zur Repräsentanz von Wissen.

So sind semantische Netze von großem Nutzen, wenn man Daten validieren will. Mit ihrer Hilfe lassen sich Relationen zwischen den interessierenden Objektklassen wie z. B. Raucher, Probanden, Nichtraucher mittels metrischer Merkmale definieren und in sogenannte probabilistische Edits überführen.

Rahmen sind ein effizientes Hilfsmittel, das Wissen zu strukturieren und die Komplexität (= Anzahl der Wissens-elemente) in einer Wissensbank zu reduzieren. Dies kann durch If-added- bzw. durch If-needed-Klauseln unterstützt werden.

Das Zusammenspiel von Datendiktionär (inkl. Dimensions- und Genauigkeitsangaben), Rahmen und Tableaus (control panel bzw. Auftragszettel (Modellierungsnotizblock)) ist aus *statistischer Sicht* noch längst nicht ausreichend genug erforscht.

### 4. Personengruppe

Die Rolle des Benutzers (user) von XPS ist noch längst nicht ausreichend formalisiert und via Benutzerprofil in ein wissensbasiertes System übertragen. Erste Ansätze zeigen, daß es wichtig ist, »trennscharfe« Klassifikationsmerkmale zu finden und möglichst früh während einer Konsultation abzufragen. Dies Problem verschwindet auch dann nicht, wenn der Entwickler zugleich Benutzer ist, da dann das Anwendungsproblem zu ermitteln ist.

## 5. Stand der Entwicklung

Es überrascht auf den ersten Blick, daß HAUX 128 Literaturstellen benennt und die Anzahl existierender Expertensysteme, die kommerziell vertrieben werden, sich auf 2 reduziert (CADEMO, GLIMPSE). Diese Relation dürfte typisch sein für den Quotienten aus prototypischen Entwicklungen und praktisch einsatzfähigen Systemen. Ein Erfahrungssatz besagt, daß man ohne Experten keine Expertensysteme entwickeln kann. Wie wir gesehen haben, kennzeichnen den Statistikexperten u. a.

- Anwendungserfahrung und
- Methodenkenntnisse.

Es ist zu vermuten, daß in vielen Fällen Defizite in der Anwendungserfahrung für den fehlenden Schritt vom Prototyp zum einsatzfähigen XPS maßgeblich waren. Es ist nicht überraschend, daß hinter den genannten beiden Systemen Forschergruppen mit viel Erfahrung bei der Benutzerberatung stehen.

## 6. Zur Konstruktion statistischer Expertensysteme

Es ist verwunderlich, daß Software-Entwicklungen im Bereich der Statistik im allgemeinen wenig Aufmerksamkeit erregen, wohl aber die Entwicklung von Expertensystemen allein schon als beachtlich gilt.

Um so wichtiger erscheint es, daß HAUX auf die Qualität statistischer Expertensysteme, Datentypen und Datenstrukturen sowie Wissensrepräsentationsformen aufmerksam macht. Um ein Gefühl von der Problematik zu bekommen, sollten Anwender statistische Expertensysteme mit der Unbefangenheit und Akribie prüfen, mit der man Software zu prüfen pflegt! Dies beginnt bei den elementaren Datentypen und endet bei der Benutzeroberfläche.

Die Datentypen integer, real, boolean und character sind unzureichend und müßten aus Sicht des Statistikers um cardinal, ordinal und classified ergänzt werden. Damit käme dem *Datendiktionär* eine viel wichtigere Rolle zu. Einträge wie Name, Typ und Länge usw. sind unzureichend. Ergänzungen um Dimension (= Maßeinheit), Skala, Rolle, Art und Querverweise sind unabdingbar, um wichtige Aspekte des Problemhintergrunds zu erfassen und eine solide *Datenvalidierung* zu gewährleisten.

Die *Datenmodellierung* muß auf die speziellen Bedürfnisse der Statistiker abgestellt werden. So werden immer wieder mehrdimensionale Tabellen redundant und semantisch aussagenarm als rechteckige Datenmatrizen abgebildet. So z. B. bei Interviews mit Probanden über deren Rauchgewohnheiten. Statt der Relation in ONF Fragebogen (*Proband Nr.*, Größe, Alter, Rauchgewohnheiten) bietet sich die Aufspaltung in 3NF-Relationen wie

- Proband (Proband Nr., Größe, Alter)
- Raucher (Proband Nr., Zigarren-Anzahl, Rauch-Dauer)
- Nichtraucher (Proband Nr., Rauch-Versuch)

an.

Die *Transparenz des Modellierungsprozesses* bei der Lösung eines statistischen Problems scheint bei vielen Expertensystemen unzureichend gelöst, wie das Auftauchen von Kontroll-Tableaus (= control-panels) zeigt, die auf einen Blick die bisher getroffenen Annahmen repräsentieren und es ermöglichen, Änderungen komfortabel durchzuführen. Die Erklärungs-komponente mit ihrer sequentiellen Vorgehensweise ist dazu nicht geeignet.

Schließlich sei noch ein Wort erlaubt zu der Funktionalität von XPS und der darin verborgenen Komplexität des Problemlösungsprozesses. Dies wird an dem von HAUX angeführten Beispiel deutlich. Allein schon die Verifikation der Annahme »Kriterium normalverteilt« ist methodisch nicht unproblematisch, da die zur Verfügung stehenden Tests geringe Schärfe haben und in XPS nach Meinung des Verfassers der Problemhintergrund

kaum erschlossen werden dürfte, um Verteilungshypothesen abzuleiten. Selbst wenn man diese Hürde genommen haben sollte, ist bei der Inferenz zu beachten, daß möglicherweise Folgen von Tests auf demselben Datenmaterial durchgeführt wurden. In vielen Systemen ist dann die Kontrolle des Testniveaus verlorengegangen.

HAUX hat angedeutet, wie es mit der Entwicklung wissensbasierter Systeme in der Statistik weitergehen kann. Zusammenfassend läßt sich sagen, daß Forschungen im Bereich der Daten- und Wissensstrukturierung, der Modellierungsunterstützung der Inferenz und der objektiven Leistungsüberprüfung, z. B. mittels Turingtest, nötig sind. Mehr Aufmerksamkeit muß dem Erschließen und Verallgemeinern von Problemhintergründen bei Beratungssystemen gewidmet werden.

Wenn in kurzer Zeit auch nicht alle – vielleicht zu hoch gesteckten – Ziele erreicht werden bzw. künftig erreicht werden können, so ergeben sich durch derartige Entwicklungsaufgaben neue interessante Einsichten in das Zusammenspiel von Angewandter Statistik und Software-Technik.

## Literatur

- NAEVE, P., J. STEINECKER, H. SCHWAN (1988): How SETUP was set up, Disk. Arb. Nr. 183/1988, Universität Bielefeld.  
 RICHTER, M. M. (1986). Architecture and Applications of Expert Systems. In: [62], 9–24.

Anschrift des Autors: Prof. Dr. H.-J. Lenz, FU Berlin, Inst. f. Quantitative Ökonomik u. Statistik, Corrensplatz 2, 1000 Berlin 33

## I. W. Molenaar

### A missing link

When HAUX writes in selection 2.2 »es wird ein konkretes Problem beschrieben«, he evidently thinks of the description of a problem by the user of the expert system. In section 2.4 he tells that the system uses the knowledge in its knowledge bank for problem solving for data from the data bank. I have argued in MOLENAAR (1988a, 1988b) that there may occur a »missing link« between these two stages: the user's problem description must be complete enough for transmission to the system, and moreover formulated in a terminology that matches that of the system.

This may not be problematic in repetitive use of the system for very similar problems: I can imagine a sequence of problems in a field like oil drilling or car maintenance in which information on a fixed and limited number of aspects of the problem is routinely available at the time when the system is consulted. Whether this favorable situation also holds for medical diagnosis decisions is more doubtful: it probably will for a specialized task like inspection of cardiograms or X-ray scans of the lungs, but it will probably not hold for general practitioners or for psychiatrists, who are faced with a large between-patient variance in the number and specificity of both the complaints and the available data.

When one turns to statistical expert systems, I am almost sure that there will be no sequence of similar problems in which decisions are based on similar sets of available information. HAND (1985) has clearly argued that a major task of a statistical consultant is to find out how the data were collected and what the major research goal is. This generally leads to an individualized dialogue in which clarification hopefully occurs. I fail to see how the client could successfully describe the problem by just answering to the pre-programmed

questions of the system. Most of my clients are not just uncertain about the right solution for their problem, they are even more uncertain about the exact formulation of their problem.

### Uncertain rules and facts

In section 2.5 the use of certainty factors for facts and rules is called a refinement of the formal representation via production rules. This topic briefly returns in section 5.5. It is clear that the impact of an expert system is substantially widened when it can be applied to »semi-facts« like »I am 70 % certain that there are Hepatitis A IgM-antibodies« or »semi-rules« like »95 % of the patients who have such antibodies suffer from Hepatitis A«. Note that the first statement probably refers to a degree of subjective belief, whereas in the second statement 95 % has the customary frequency interpretation.

I have problems with the use of fuzzy logic and certainty factor calculus as explained in e.g. HOLSAPPLE and WHINSTON (1986) or PARSAY and CHIGNELL (1988). Like many other authors in this area, they begin by correctly warning the reader that the use of probabilities in expert systems has its problems. The set of basic events that the user has in mind may contain overlaps, accurate assessments of some prior probabilities may not be available, combination rules for probabilities are complicated if there is dependence between the events involved (which is often plausible). Next, however, fuzzy logic and certainty algebra are introduced as a solution to these problems. I fail to see why the same problems would not invalidate the use of these alternative expressions for uncertainty: it may be easier to hide them, but that is not very helpful. Assessment and combination of probabilities is difficult, but remains so when one replaces probabilities by certainty factors or by verbal labels of fuzziness.

### Decisions about the appropriateness of a method

At the end of section 4.1 HAUX writes »... sollen solche Systeme in verstärktem Maße darauf achten, daß das verwendete statistische Verfahren oder das verwendete statistische Modell für das zu untersuchende Problem angemessen ist«. I could not agree more, but I see a major problem. This was already a central issue at a discussion during COMPSTAT 1978 in Leiden between the representatives of five major statistical package distributors and their angry users. In so far as the data themselves contain information that is relevant for such appropriateness decisions, there has been some progress: the user gets more warnings about things like small expected values in a contingency table, multicollinearity in a regression problem, skewness of distributions.

The data themselves, however, are not enough for the appropriateness decision. The sample sizes may e.g. be too small for sound decisions on skewness or equality of variances. Even worse, the data can never answer the key question whether they have been collected by careful measurement and proper random sampling, or more generally whether it is appropriate to consider them as random samples of the intended variables from the populations to which one desires to generalize the conclusions of the study.

A clear example of a too hasty inference from the data alone occurs in HAUX's Figure 5 illustrating REX, where an »unduly skew« distribution of  $y$  automatically leads to the advice to use  $\log y$  instead as the regression criterion. This bypasses both the question why the variable has a high skewness coefficient (it may be more appropriate to omit a few outliers or to correct a few wrong data entries of  $y$ -values), and the question whether the transformation improves the linearity of the criterion as a function of the predictor(s): a linear relation for the conditional expectation of  $y$  given  $x$  may very well go together with a skew overall distribution of  $y$ , and in that case  $\log y$  may be a rather non-linear function of  $x$ .

### GLIMPSE, PANOS, REX

The clear introduction to these three systems leaves me with only two minor remarks. HAUX says in 4.5 »REX dient, wie GLIMPSE, der explorativen Datenanalyse«. This is correct in so far as both offer the possibility to fit various models to the same data. My own view, however, is that REX stays closer to Tukey's ideas of EDA, whereas GLIMPSE considers classes of parametric models exactly specified by their link function and error function, in which the likelihood principle is used for parameter estimation and goodness of fit tests.

As regards PANOS, I read in 4.4 »der Benutzer – vorausgesetzt wird eine Fachperson«, but a few lines later the system is called »autoritarian«, making me think that it is »system guided« with limited room for initiatives of the user. I would expect an authoritarian system for a novice and a more user guided system for a »Fachperson«. Did I misunderstand the point or were there special reasons for the authoritarian structure?

### Three cheers for section 5

It was a real pleasure, and a shock of recognition, to read HAUX's section 5. His criticism of the current state of the art and his warnings against unduly optimistic science fiction are well founded. The emergence of systems based on too simplified statistical rules would run fully against the aim of diminishing abuse of statistical methods, often quoted as one of the main reasons for building the systems. In MOLENAAR (1989) I try to sketch the more general problem that the scientific world has not yet found quality control mechanisms for software of the same effectivity with which we screen scientific books and papers, or maintain quality levels for goods like cars or household electronics. Papers like the one commented on here are a valuable contribution to this quality discussion, but we probably need more activity of scientific gremia too, perhaps even legislation.

### Time for a freeze

In his final section HAUX mentions a desirable side effect of the development of statistical expert systems: they force the statisticians to formal representation of their knowledge and to propose suitable inference mechanisms for their problem solving activities. This justification for research in this area has been given by many, among which myself. In the past year, however, my skepticism has increased whether the current emphasis on building systems is the most fruitful road to solving our problems. Not only do I perceive very little successful activity in the area of formal representation of statistical knowledge and statistical inference rules. Even worse is the almost complete absence of field tests of the systems by representatives of the class of intended users. I would argue in favor of freezing, or at least slowing down, the building and marketing of systems until the scientific community has established more clearly what knowledge should go into statistical expert systems, how it can be represented and used by the systems without the combinatorial explosion sketched by STREITBERG (1988), and what results from a systematic study of the use of the existing systems by a well chosen sample of users.

### References

- HAND, D. J. (1985): Choice of statistical technique, IASC Selected Papers 45th ISI Session, 285–300.  
HAUX, R. (1989): Statistische Expertensysteme. *Biometrie und Informatik in Medizin und Biologie* 1-2, 3.

- HOLSAPPLE, C. W., A. B. WHINSTON (1986): Manager's guide to expert systems using GURU, Homewood (Ill): Dow Jones-Irwin.
- MOLENAAR, I. W. (1988a): Statistical consultants and statistical expert systems. In: D. F. EDWARDS and N. E. RAUN (eds), COMPSTAT 1988 Proceedings, 187-192, Heidelberg: Physica Verlag.
- MOLENAAR, I. W. (1988b): To exist or not to exist. Statistical Software Newsletter **14**, 127-128.
- MOLENAAR, I. W. (1989): Producing, purchasing and evaluating statistical software: the role of intellectual and commercial challenges, Invited address at SOFTSTAT '89, Heidelberg.
- PARSAY, K., M. CHIGNELL (1988): Expert systems for experts, New York: Wiley.
- STREITBERG (1988): On the non-existence of expert systems, Critical remarks on artificial intelligence in statistics. Statistical Software Newsletter **14**, 55-62, with discussion, 63-74.

Author's address: Ivo W. Molenaar, Vakgroep S & M FPPSW, Oude Boteringestr. 23, 9712 GC Groningen, Netherlands

## P. Naeve

. . . Der Riese las: »Sieben auf einen Streich«, meinte, das wären Menschen gewesen, die der Schneider erschlagen hätte, und kriegte ein wenig Respekt vor dem kleinen Kerl. Doch wollte er ihn erst prüfen, nahm einen Stein in die Hand und drückte ihn zusammen, daß das Wasser heraustropfte. »Das mach mir nach«, sprach der Riese, »wenn du Stärke hast«. »Ist's weiter nichts?«, sagte das Schneiderlein, »das ist bei unsereinem Spielwerk«, griff in die Tasche, holte den weichen Käse und drückte ihn, daß der Saft herauslief. . . . [3, S. 144]

Dieser Anfang mag für Laien und Experten gleichermaßen befremdlich sein. Aber erinnern wir uns, in Märchen steckt viel Wahrheit und Weisheit. Und sie liefern uns schöne »Bilder«, mit deren Hilfe sich auch schmerzhaft Wahrheiten leichter aussprechen lassen.

Doch zuerst muß die Tapferkeit von Reinhold Haux gepriesen werden, der weit mehr als sieben Veröffentlichungen durch einen Streich in einem Übersichtsartikel zusammenfaßte. Es ist jetzt leicht, das Ergebnis zu kommentieren und gar zu kritisieren. Ich bin froh, daß die Herausgeber der Zeitschrift nicht die Reihenfolge Übersichtsartikel durch Naeve und Kommentar von Haux gewählt haben.

Wie wäre es mit einer Gleichsetzung von Expertensystem und dem Stein des Märchens? Scheinen doch die Expertensysteme dem Laien bei erster »Berührung« hart und verschlossen wie ein Stein – oft mit polierter Oberfläche. Hat denn nun aber der Stein wirklich einen Gehalt, den ein Riese herauspressen könnte (hier natürlich in Form eines gehaltvollen Artikels)? Oder ist er lediglich taubes Gestein? Die andere mögliche Gleichsetzung von weichem Käse und Expertensystem läßt sofort Erinnerungen an Goethe hochkommen:

Getretener Quark  
Wird breit, nicht stark. –  
Schlägst du ihn aber mit Gewalt  
In feste Form, er nimmt Gestalt.  
Dergleichen Steine wirst du kennen,  
Europäer Pisé sie nennen. [2, S. 51]

Quark oder gibt es sie vielleicht gar nicht [9]? Wie sollte man diese Frage klären? Man könnte zum Beispiel nach einer Definition suchen, die in operationalisierbarer Form ein statistisches Expertensystem beschreibt, und damit die angebotenen Kandidaten konfrontieren. Man könnte sich eine Liste von mit Hilfe eines statistischen Expertensystems zu lösenden Aufgaben anlegen und prüfen, ob die angepriesenen Kandidaten die Aufgaben tatsächlich lösen. Arbeitet man sich mit diesen beiden Vorstellungen im Hinterkopf durch

den Berg von Veröffentlichungen – trotz seines Fleißes hat Reinhold Haux ihn für seinen Artikel nicht abtragen können –, so gelangt man als erstes zu der Erkenntnis, daß das ganze Gebiet durch eine ungeheuer schwammige Begrifflichkeit gekennzeichnet ist. Leider ist auch der Artikel von Haux an einigen Stellen davon nicht unberührt geblieben.

Wenn man liest: »Wir wollen eine Maschine . . . genau dann als Expertensystem bezeichnen, wenn sie Funktionen, die für einen Experten spezifisch sind, übernehmen kann . . . Unter (spezifischen) Funktionen . . . verstehen wir diejenigen Tätigkeiten, die, außer von Expertensystemen, nur von Experten übernommen werden können.«, dann ist der Schritt, alles und jedes als Expertensystem zu bezeichnen, nur noch eine Kleinigkeit. Dann fällt es auch nicht mehr schwer, Streitbergs These zu widerlegen.

Die den Bereich Expertensysteme kennzeichnende Vagheit der Begriffe gewinnt noch an Gefährlichkeit durch die Neigung, Schwierigkeiten durch Zuweisung an alte oder neue Begriffe zu lösen. Stellvertretend sei hier auf den Begriff Schnittstelle verwiesen. In der Informatik leistet er für die Beschreibung (Hard- und Software) technischer Koppelung von Systemen gute Dienste, wenn man aber im Zusammenhang mit Expertensystemen diesen Begriff verwendet, dann darf man nicht übersehen, daß man in der Regel mehr als eine rein technische Interpretation im Sinne hat. Über die »Schnittstelle« werden dann nämlich »Weltsichten« verbunden (die des Expertensystems und die des Benutzers z. B.). Und diese Problematik ist alles andere als gelöst. Man hat auch eher den Eindruck, daß bei der beliebten Verwendung des Begriffes »Schnittstelle« durch den impliziten Verweis auf die technische (bewältigte) Seite die Probleme verdeckt werden (sollen).

Da die meisten Menschen sich einem Expertensystem wohl in der Rolle eines (potentiellen) Benutzers nähern, einige wenige sich als abzubildende Experten sehen, wird leicht die Person im Hintergrund übersehen. Bei Haux trägt sie den Titel Wissensverwalter, in der Literatur ist sie besser als knowledge engineer bekannt. Listet man alle Tätigkeiten auf, die ihm in der Literatur zugewiesen werden, und alle Schwierigkeiten, deren Bewältigung ihm übertragen wird, so drängt sich der Eindruck einer omnipotenten Persönlichkeit auf. Er soll nicht nur Experte in dem Einsatzgebiet des Expertensystems sein, er muß Informatiker sein, um die Systeme »programmieren« zu können, und er muß Psychologe sein, um das Wissen anderer Experten aus diesen »herauszuziehen«. Hat sich schon einmal jemand gefragt, woher auf einmal alle diese »Wissensverwalter« kommen sollen? Wo wird man seriös zum knowledge engineer ausgebildet?

Die wohl wichtigste Frage, die man sich stellen muß, ehe man sich »statistischen Expertensystemen« zuwendet, ist aber: Was macht eigentlich einen statistischen Experten aus? Natürlich gibt es – schon lange vor dem Aufkommen von Expertensystemen – eine umfangreiche Literatur zu diesem Thema. Man lese die »presidential addresses« im Journal of the Royal Statistical Society oder aber – mehr im Zusammenhang mit Expertensystemen – im von W. Gale [1] herausgegebenen Sammelband. Aber man gewinnt nicht den Eindruck, daß die »Erbauer« von sogenannten statistischen Expertensystemen sich mit dieser Literatur auseinandersetzen oder gesetzt haben. Da die Diskussion auch für die Kritiker von Expertensystemen von Nutzen ist, sei hier der Versuch einer Beschreibung eines statistischen Experten gemacht:

#### **Vorbemerkung:**

Die folgende Definition eines »statistischen Experten« wird vielleicht von vielen Leuten als die den Stand der Überlegungen beschreibende akzeptiert werden.

An »Expert System« is regarded as the embodiment within a computer of a knowledge-based component from an expert skill, in such a form that the system can offer intelligent advice or take an intelligent decision about a processing function. A desirable additional characteristic which many would consider fundamental, is the capability of the system, on

demand, to justify its own line of reasoning in a manner directly intelligible to the enquirer. The style adopted to attain these characteristics is rule-based programming (Hand [4]).

Da man ähnliche Definitionsversuche auch in anderen Wissenschaftsbereichen antrifft, sieht es auf den ersten Blick so aus, als ob ein statistisches Expertensystem sich von einem medizinischen Expertensystem im Prinzip nur durch ein anderes Prefix unterscheidet. (Ist das zunehmende Angebot von Expert System Shells nicht eine Bestätigung dieser These?) Daß dem nicht so ist, zeigt das folgende Zitat:

What is it what we're trying to sell? Is it our considerable collection of arithmetic techniques and computer programs? I think not. . . . In my opinion, what we have to sell that is worth the effort is the scientific method, or the logic of experimentation. . . . Clearly the whole scientific world may not yet be ready to admit that statisticians have become the custodiens of the scientific method, but here are nevertheless some cogent arguments in support of this idea: Every good experimental statistician I know is deeply interested in the logic of experimentation. Physicist, on the other hand, are mainly interested in physics, and chemists in chemistry. While they believe in the scientific method, and use it, they usually believe that they are so imbued with it that they seldom need to think about it. Some philosophers are interested in the scientific method, but few of them actually interact with the world of experimentation. So we are inheriting the scientific method because we care about it, we write about it, and we do something about it, and no one else seems to do all these things (Hooke [5]).

#### Komponenten statistischer Expertise:

Zwischen dem Zitat von HOOKE in den Vorbemerkungen und diesem Abschnitt wäre eigentlich der Platz für eine längere Abhandlung, was denn nun genau für praktisches statistisches Tun die Verpflichtung zur Beachtung der »scientific method« bedeutet. Hier sei nur so viel angedeutet, daß gerade die beiden ersten im folgenden aufgeführten Komponenten Konsequenzen der »logic of experimentation« sind (und erst recht die Ausführungen des nächsten Abschnitts). Statistische Expertise läßt sich unter den drei Überschriften klassifizieren:

**Daten-Expertise:** Analogieschluß, z. B. der saisonale Effekt in dieser Zeitreihe wird vermutlich dem einer früher analysierten Zeitreihe in einem ähnlichen Anwendungsbereich gleichen.

Anhand von Datenbeschränkungen, z. B. wenn die Datenpunkte jährlich erhoben wurden, lassen sich keine Informationen über Saisonkomponenten gewinnen.  
etc.

**Modell-Expertise:** Modellvoraussetzungen, z. B. ARMA-Modelle, setzen Stationarität voraus.

Modellkonsequenzen, z. B. die Annahme eines additiven Modells, lassen eine logarithmische Transformation wenig geeignet erscheinen.  
etc.

**Methoden-Expertise:** Methodenvoraussetzung, z. B. positive Werte bei einer logarithmischen Transformation, Normalverteilung der Daten bei gewissen Tests.

Konsequenzen, z. B. Änderung der Verteilung nach einer »varianzstabilisierenden Transformation«.

#### Der Prozeß des statistischen Schließens:

Eines der wichtigsten Konzepte in der Statistik ist der Begriff des statistischen Modells. Alle Probleme werden explizit oder implizit vor dem Hintergrund eines statistischen Modells

gelöst (bzw. zu lösen versucht). Der Begriff statistisches Modell ist dabei sehr weit gefaßt, d. h., er umfaßt quantitative und qualitative Modelle.

Z. B. seien Beobachtungen  $x_1, \dots, x_n$  gegeben. Dann sind denkbare Modelle:

- i) beschreibende Modelle i.e. zusammenfassende Charakterisierung der Daten, z. B.  $\bar{x}, s^2$ , Häufigkeitstabellen usw.
- ii) Modell für Skalenniveau: Vorstellung über Meßvorschrift bzw. Prozeß der Datengenerierung
- iii) Modell für Zusammenhang »Stichprobe« und »Gesamtheit«
- iv) Verteilungsmodell  $\{F_\theta(x)\}_{\theta \in \Theta}$  bzw.  $F_{\theta_0}(x)$ .
- v) qualitatives Modell für Grundgesamtheit: unimodal, symmetrisch, heavy-tail usw.
- vi) etc.

Statistisches Schließen läßt sich nun beschreiben als das Suchen nach einem geeigneten Modell. Dieser Suchprozeß verläuft nun in zwei Phasen:

- i) Intra-level-search
- ii) Inter-level-search

Ein Beispiel möge an dieser Stelle den Zugang erleichtern. Betrachte die Modellierungsaufgabe für eine Zeitreihe  $x_t$ . Der Suchprozeß habe sich bisher so entwickelt, daß man sich dem Box-Jenkins-Ansatz genähert hat. Formaler läßt sich die Situation wie folgt fassen:

Level  $i$  : AR-Modell, MA-Modell, ARMA-Modell, ?-Modell  
 $\downarrow$   
 Level  $i+1$  : AR(1), AR(2), . . . AR( $q$ ), AR(?).

Die Phase des Intra-level-search berücksichtigt nun die Tatsache, daß der Statistiker mehr als ein Modell zur Zeit verfolgt. Hier im Beispiel auf dem Level  $i$  AR-Modell, MA-Modell und ARMA-Modell. Außerdem ist ihm zu jeder Zeit bewußt, daß es auch noch weitere von ihm noch nicht spezifizierte Modelle gibt. Hier mit ?-Modell bezeichnet, man denke z. B. an ARIMA-Modelle. Den »Intra-level-search« kann man formal wie folgt fassen:

Auf jedem Level gibt es Modelle

$$M_1, \dots, M_n \quad n = f(\text{level}); M_n \text{ unspezif. Rest}$$

mit Plausibilitäten  $P(M_i) \geq 0, i = 1, \dots, n$ . Es gelte  $\sum P(M_i) = 1$ .

Weiterhin sei auf jedem Level ein Satz von Statistiken  $S_j, j = 1, \dots, m; m = g(\text{level})$  gegeben. Für diese Statistiken seien die bedingten Wahrscheinlichkeitsverteilungen  $P(S_j|M_i), i = 1, \dots, n-1$  gegeben.

Für das Folgende sei ein Diskretisierungsprozeß für die Statistiken  $S_j$  unterstellt, so daß man die folgende Matrix aufstellen kann für jede Statistik:

$M \setminus S$	$S_j = s_{j1}$	$S_j = s_{j2}$	...	$S_j = s_{jm}$
$M_1$	$P(S_j = s_{jk}   M_i)$			
$M_2$				
$\cdot$				
$\cdot$				
$M_{n1}$				
$M_n$				

Mit Hilfe der Statistik  $S_j$  kann nun der Intra-level-Search, d. h. ein Schritt dieses Suchprozesses, durchgeführt werden. Dieser Schritt besteht in einem Update-Prozeß für die Plausibilitäten der betrachteten Modelle  $M_1, \dots, M_{n-1}, M_n$

$$P(M_i) = P(M_i) \frac{P(S_j = s_{jk} | M_i)}{P(S_j = s_{jk})}$$

wobei

$$P(S_j = s_{jk}) = \sum_{i=1}^n P(S_j = s_{jk} | M_i) P(M_i)$$

(Ein praktisches Problem liegt sicher in der Spezifikation der von  $M_n$  – dem »Restmodell« – abhängigen Größen, konzeptionell ist diese Frage aber zur Zeit nicht von Belang.)

Man kann das geschilderte Verfahren als eine Verallgemeinerung der »Differentialdiagnose« bezeichnen. Aber dieser Bezug gilt nur für einen Schritt, denn es kann nach jedem Schritt zu einer Veränderung der Kandidatenmenge  $\{M_1, \dots, M_{n-1}, M_n\}$  kommen, da Modelle  $M_i$   $i=1, \dots, n-1$  mit  $M_n$  verschmolzen werden bzw. neue Modelle  $M_i$  aus dem »Restmodell«  $M_n$  abgespalten werden.

Die in jedem Schritt benötigte Statistik  $S_j$  kann gemäß ihres Diskriminationsindex  $\Delta(S_j)$  ausgewählt werden.

$$\Delta(S_j) = \frac{\Delta'(S_j)W(S_j)}{\sum_{i=1}^n \Delta'(S_i)W(S_i)}$$

$$\Delta'(S_j) = \sum_{l,k=1}^n \Delta_{l,k}(S_j)P(M_l)P(M_k)$$

$$\Delta_{l,k}(S_j) = \sum_s |P(S_j = s | M_l)P(M_l) - P(S_j = s | M_k)P(M_k)|$$

Dabei sind die  $W(S_j)$  vom Problembereich her bestimmte Gewichte. Die Menge der zur Verfügung stehenden Statistiken  $S_j$  kann sich nach jedem Schritt ( $\rightarrow$  siehe Änderung der Modelle) verändern. Dies wirkt sich aber nur als Hinzufügen bzw. Wegnehmen einer Matrix ( $P(S_j | M_i)$ ) aus. Modelle, die hinzukommen, fügen eine Zeile zu den Matrizen, die in  $M_n$  verschwindenden Modelle führen zu einer Streichung von Zeilen. Die Zeile für  $M_n$  wird in jedem Fall modifiziert.

Unsere Bemühungen haben bisher ergeben:

- i) Mechanismus zur Fortschreibung von Modellplausibilitäten
- ii) Möglichkeit der Geburt und des Todes (Schlafens) von Modellen (Kreativität des Experten).

Den Vorgang der Intra-level-Suche könnte man auch als sideward-chaining bezeichnen. (Dieses Wort wurde von Prof. Streitberg geprägt.)

Irgendwann wird plötzlich (kreativ) bzw. als Abschluß eines Reifeprozesses (Modellplausibilität nahezu Ein-Punkt-Verteilung) zu einem anderen Level übergegangen. Dies bedeutet dann einen neuen Satz von Modellen und Statistiken. Auf diesem neuen Level beginnt dann wieder ein Intra-level-search der beschriebenen Art. (Im Beispiel verdichtete sich die Plausibilität auf dem AR-Modell, so daß es zum Übergang auf einen Level differenzierter AR-Modelle kam. Es wären auch andere Überlegungen denkbar, z. B. differenzierte MA-Modelle.)

Der Umgang mit den Statistiken auf jeder Ebene erfolgt dann in einer Form, die man als self guarding method bezeichnen kann.

Problem : Transformation von Zeitreihen  
 Methode : Logarithmus-Transformation  
 Guard : if all Data >0 then logarithmic transformation possible  
 Wissen : if data are of type age then data >0  
 Fakt : time series data are age of person

◊ Man kann Methode anwenden bzw. Methode »läßt« sich anwenden.

Wenn man diesen Ausführungen im großen und ganzen zu folgen bereit ist, stimmt man auch der Aussage zu: »Es gibt (zur Zeit) keine statistischen Expertensysteme. Was es gibt, sind Ergebnisse kleiner Fingerübungen auf extrem eingeschränkten »systematic domains.« [11, S. 174]

Hier wäre nun die richtige Stelle für den Einwand, daß die Situation ja nicht zuletzt deswegen so sei wie sie sei, weil es bisher an der Verfügbarkeit der geeigneten Werkzeuge zur Konstruktion von Expertensystemen gefehlt habe.

Aber wem die Aufgabe, das offenbarte Expertentum mit Hilfe von Programmiersprachen wie PROLOG, LISP usw. umzusetzen, – die sich den oben erwähnten neuen Formen der Wissenspräsentation und Inferenz verpflichtet fühlen – zu schweißtreibend erscheint, dem kann geholfen werden. Shell ist die Zauberformel.

Eine Shell ist eine Schale ohne Inhalt. Man stelle sich die von HAUX gezeigte Struktur des Expertensystems ohne jeglichen Bezug zur Problemdomäne vor. Alles, was der Experte – oder der Knowledge Engineer (Wissensverwalter) – noch zu tun hat, ist die Expertise in die Schale zu füllen. (Kommt da Erinnerung an Frankenstein hoch?)

Aber es gibt bei näherem Hinsehen nach wie vor Probleme – vielleicht sogar noch größere.

### 1. Interaktivität

Die statistische Software hat sich seit den Zeiten der BMD-Programme doch erheblich verbessert. Interaktives Arbeiten ist kein Wunschtraum mehr.

Die Shells sind nun so konzipiert, daß sie nicht nur das letzte, sondern überhaupt das Wort haben wollen. Zur »Statistischen Software« – die ja ein Hort von prozeduralem Wissen ist – haben sie ein Master-slave-Verhältnis.

```
master: Shell
slave: Statistikpaket, z. B. P-STAT
```

Was uns dies kostet, läßt sich bei der Umsetzung eines explorativen Spiels in interaktiver Manier mit P-STAT zeigen. Die Daten befinden sich in library 11 in file f1 als Variable v1 . . . vn.

#### i) P-STAT direkt

```
pstat
find f1, library 11
eda f1;

median.polish v1, row vi, col vj;
var v2;
steps 4;
etc.
```

## ii) über eine Shell

```

overhead 1:  batch $
              find f1, library l1 $
              eda f1; out outfile;
median.polish v1, row vi, col vj

overhead 2:  list outfile, pr outfile 1 (.data.only)$
              end $

»overhead 1«
median.polish v2, row vi, col vj

»overhead 2«
etc.

```

## 2. Interaktivität und Graphik

Zunehmend erlauben statistische Systeme interaktives Umgehen mit Graphik. Ungelöste Probleme sind dabei

- Interaktivität wie eben besprochen
- kontrolliertes Beschreiben, was man in der Graphik sieht und macht, d. h., das Expertensystem muß dem Benutzer über die Schulter schauen können. Ein Fall von Tukeys »cognistics«.

3. Die Shells sind genau betrachtet doch nicht leer. Im allgemeinen kommt den Shells die Bezeichnung »hybrides System« zu, d. h., sie unterstützen

- verschiedene Formen der Wissensrepräsentation
- verschiedene Formen der Inferenz.

Damit entsteht aber das Problem des Metawissens. Wie geht das System damit um,

- wie wählt es zwischen verschiedenen Interferenzmethoden aus, z. B. erst forward, dann backward, immer zuerst Tiefe, dann Breite usw.
- welche Einflußmöglichkeiten hat die Wissensrepräsentation: frame versus production rule.

Das Paradigma der Expertensystem-Shell ist nicht offengelegt – geschweige denn offen für Modifikation durch den Benutzer.

Über ein größeres Projekt (mehr von seinen Schwierigkeiten), ein statistisches Expertensystem auf der Basis einer verfügbaren Shell und mit verfügbarer Statistiksoftware zu erstellen, wird vom Verfasser an anderer Stelle [8] berichtet.

Ich unterstütze die kritischen Anmerkungen von Reinhold HAUX. Es ist aber wohl deutlich geworden, daß ich mir zu vielen Stellen ein noch viel schärferes, kritisches Hinsehen gewünscht hätte. Die größte Gefahr, die der statistischen Kultur durch sogenannte statistische Expertensysteme zur Zeit droht, läßt sich an einer Definition des Begriffes »intelligent« festmachen. Im Merrian-Webster [10] kann man lesen:  
intelligent . . . 3. able to perform computer functions . . .

Die Gefahr heißt Reduktionismus. Statistische Expertise ist das, was sich in einem Expertensystem abbilden läßt. Das, was beim »Hineingießen« in eine Shell vorbeiläuft, ist eben keine statistische Expertise. Im Lichte der Reduktionsgefahr ist die eingangs erwähnte Begriffsbestimmung eines statistischen Expertensystems fast schon wie die Anwendung der Merrian-Webster-Definition.

Um auf die Frage nach Stein oder Quark zurückzukommen: Zur Zeit schaffen die »Statistischen Expertensysteme« – auch wenn es noch gar keine gibt – mehr Probleme als sie lösen, zumindest wenn man sich der Materie seriös nähert.

Es gibt auch Statistiker, die verkünden, »es gibt heute keine, es gibt auch morgen keine, weil es keine geben kann«, indem sie über Komplexitätstheorie argumentieren. Aber auch mit denen bin ich mir über die folgenden Argumente einig, warum wir Statistiker uns dennoch mit Expertensystemen beschäftigen sollten:

- Sonst tun es andere (schlechtere Statistiker); dies ist ein Notwehrargument.
- Zur Gewinnung einer »wissenschaftlich« vertretbaren Meinung, wir müssen uns, ob pro oder contra, von der Marketing-verseuchten EDV-(Informatiker-)Szene absetzen. [Mensch sei seriös dagegen.]
- Wir lernen verstehen, was wir tun, wir haben mit den geschilderten Wissensrepräsentationen und Inferenzverfahren sprachliche Konstrukte, um – fern vom Computer – unser statistisches Tun zu analysieren.
- Wir gewinnen neue statistische Problemstellungen und hoffentlich auch Lösungen, z. B. Tukeys cognostics problem.

Fazit: Sicher ist nur, daß der Statistiker Nutzen aus der Beschäftigung mit statistischen Expertensystemen zieht, für die Klienten oder die statistische Kultur ist dies fraglich – über Kommerz schweigt man lieber.

## References

- [1] GALE, W. A. (ed.) (1986): *Artificial Intelligence & Statistics*, Addison-Wesley, Reading.
- [2] Goethes Werke, Weimarer Ausgabe, 6. Bd., S. 131, Weimar 1888, Fotomechanischer Nachdruck, München 1987
- [3] Das tapfere Schneiderlein. In: *Kinder- und Hausmärchen*, gesammelt durch die Brüder Grimm, Winkler Verlag, München 1963
- [4] HAND, D. J. (1984): Statistical expert system design, *The Statistician*, Vol. 33.
- [5] HOOKE, R. (1980): Getting people to use statistics properly, *The American Statistician*, Vol. 34, No. 1.
- [6] NAEVE, P., D. METZING (1987): Anwendung von TWAICE in der Statistik. In: CREMERS, A. B., D.-H. BECK (eds.) *Proceedings Anwenderforum Expertensysteme*, Wuppertal.
- [7] NAEVE, P., J. STEINECKER (1987): SETUP – Statistisches Expertensystem mit TWAICE und P-STAT, Diskussionspapier Nr. 172/1987 der Fakultät für Wirtschaftswissenschaften, Universität Bielefeld.
- [8] NAEVE, P., J. STEINECKER, H. SCHWAN: How SETUP was set up, Diskussionspapier Nr. 183/1988 der Fakultät für Wirtschaftswissenschaften, Universität Bielefeld.
- [9] STREITBERG, B. (1988): On the nonexistence of expert systems. Critical remarks on artificial intelligence in statistics, *Statistical Software Newsletter*, Vol. 14, No. 2.
- [10] Webster's Ninth New Collegiate Dictionary, Merriam-Webster Inc., Springfield 1983.
- [11] WINOGRAD, T., F. FLORES (1986): *Understanding Computers and Cognition*, Ablex Publ. Corp., Norwood.

## D. Rasch

Mir hat die Übersichtsarbeit von Herrn Haux sehr gut gefallen, aber zwei Aspekte kamen mir dabei zu kurz:

1. So vielfältig wie die Aufgaben eines Statistikers sind, ebenso vielfältig können Expertensysteme der Statistik eingesetzt werden. Ich möchte mich hier nicht mit den Problemen eines solchen theoretischen Statistikers befassen, der seine Aufgabe in der Herleitung neuer Sätze und im Formulieren neuer Definitionen und Theorien sieht. Expertensysteme, die ihm hilfreich sein können, haben eine besondere Struktur und sind auch in anderen Bereichen der Mathematik einsetzbar. Vielmehr möchte ich etwas zu Expertensystemen äußern, die den beratenden Statistiker unterstützen und zu Expertensystemen, die an seine Stelle treten können. Solche Expertensysteme müssen vor allem in der Lage sein, eine im Einsatzgebiet der Statistik formulierte fachliche Fragestellung (z. B. eine Sachhypothese) in eine statistische Fragestellung (z. B. eine statistische Hypothese) zu transformieren. Dies ist der schwierigste Schritt bei der Nutzung statistischer Methoden, man könnte ihn mit der Diagnose in der Medizin vergleichen. Nur erfahrene Statistiker sind in der Lage, diesen Schritt – wir wollen ihn Modellierung nennen – in unterschiedlichen Anwendungsgebieten richtig durchzuführen. Das Problem besteht darin, daß oft mehrere statistische Modelle zur Beschreibung des Sachverhaltes und zur Lösung der Aufgabenstellung geeignet sind. Nicht selten kommt es vor, daß ein Wissenschaftler, wenn er mehrere Statistiker konsultiert, mehrere Modelle angeboten erhält. Dem muß ein entsprechendes Expertensystem Rechnung tragen, indem während des Dialoges öfter nachgefragt wird, ob der Nutzer des Systems denn wirklich im richtigen Zweig ist. Wir wollen einmal annehmen, der Nutzer hat sich im Gewirr der multiplen Mittelwertvergleiche verstrickt und kann sich nicht recht für eines der angebotenen Risiken erster Art entscheiden.

Dann wäre die Frage nützlich: »Wollen Sie wirklich Ihre 5 Behandlungen miteinander vergleichen, um festzustellen, ob wesentliche Unterschiede vorliegen, oder geht es Ihnen nicht eher darum, herauszufinden, welche Behandlung die beste (also etwa die mit dem größten Mittelwert) ist?«

Entscheidet sich der Nutzer für den zweiten Weg, gelangt er zu den Auswahlverfahren (Bechhofers Indifferenzbereichsformulierung oder Guptas Teilmengenformulierung). Offensichtlich hat er den Absprung zu diesen Verfahren an einer früheren Stelle verpaßt.

Solche Möglichkeiten entsprechen der Veränderung der Diagnose während der Therapie.

In einem zweiten Beispiel wollen wir annehmen, ein Rinderzüchter will die Milchmengenleistungen zweier Rinderrassen an Hand der Leistungen zweier großer Herden mit Tieren je einer dieser Rassen vergleichen. Ein Expertensystem, das hier den t-Test empfiehlt ohne zu fragen, ob die Tiere in beiden Herden nicht eventuell selektiert sind (dann wäre der Wilcoxon-Test zu empfehlen), dürfte den oben gestellten Anforderungen an eine adäquate Modellierung nicht genügen. Beim Vergleich von Mittelwerten ist natürlich nicht nur auf etwaige Stützung von Verteilungen (z. B. durch Selektion), sondern auch auf etwaige Zensur der Stichproben zu achten (beides geschieht z. B. im Modulkomplex »Beurteilung von Mittelwerten und Wahrscheinlichkeiten« von CADEMO).

2. Die Mehrzahl der Statistiklehrbücher behandelt die Mathematische Statistik als die Theorie der Aufbereitung von vorliegenden Daten und des Schließens von diesen Daten auf eine Grundgesamtheit. Dabei wird nach optimalen Verfahren (Schätzungen, Tests) gesucht, die zu einem gegebenen Beobachtungsumfang ein Maximum an Sicherheit garantieren. RASCH und HERRENDÖRFER (1982, 1986) haben darauf hingewiesen, daß die Mathematische Statistik als Einheit der statistischen Versuchsplanung und der statistischen Auswertungsstrategie anzusehen ist. Nur unter besonderen Umständen kann die Minimierung einer globalen Risikofunktion so vorgenommen werden, daß man getrennt den optimalen Versuchsplan und die optimale Auswertung aufsucht (siehe die in oben genannter Literatur angegebenen Trennsätze) Zumindest sollte man sich vor der Durchführung von Versuchen

und Erhebungen Gedanken machen, wie man sein Problem mit minimalem Versuchsumfang lösen kann. Das kann in vielen Fällen wie z. B. in der Regressionsanalyse mit einstellbaren Einflußgrößen durch die optimale Wahl der Meßstellen erfolgen. Bei statistischen Tests zum Mittelwertvergleich kann man nach Vorgabe der Risiken erster und zweiter Art und einer praktisch interessierenden Mindestdifferenz den minimalen Versuchsumfang ausrechnen. Tut man das nicht, erhält man immer Signifikanz, wenn man nur oft genug mißt (Signifikanz ist eine Fleißarbeit). Daraus ersieht man, daß Expertensysteme in der Statistik in der Erkenntnisgewinnung an drei Stellen einzusetzen sind

- Modellierung des Sachverhaltes durch ein statistisches Modell
- Optimale Versuchsplanung (Planung von Erhebungen)
- Versuchsauswertung.

Die von HAUX beschriebenen Systeme beziehen sich nach seiner Darstellung auf die dritte Stelle.

Das gilt aber zumindest für CADEMO nicht. CADEMO unterstützt gerade die beiden erstgenannten Schritte. Es gibt dann lediglich Hinweise auf die Methode, mit der nach dem Versuch ausgewertet werden sollte, falls der Versuch nach dem von CADEMO vorgeschlagenen Versuchsplan (einschließlich Umfang) durchgeführt werden konnte.

CADEMO wendet sich sowohl an Statistiker, als auch an die Versuchsansteller selbst. Damit sich der Statistiker nicht langweilt, kann er kurze Wege gehen und auf den Aufruf von Erklärungen unbekannter Begriffe (entnommen aus RASCH (1988)) verzichten.

Zum Schluß soll auf die Expertensysteme N von idv, SELINA (BAINES u. a. 1988) und Planung (HEISELBETZ 1985, 1986) verwiesen werden, die ebenfalls die Versuchsplanung unterstützen. Im Gegensatz zur Bemerkung von HAUX unter 4.2 steht die Tatsache, daß auch N und SELINA kommerziell vertrieben werden.

## Literatur

- BAINES, A., D. T. CLITHERO, and P. J. ZEMROCH (1988): SELINA – A conversational package for the statistical design and analysis of experiments. *Compstat* **88**, Software catalogue, Kopenhagen, 13–14.
- HEISELBETZ, C. (1985): PLANUNG – Hilfsmittel für die Versuchsplanung. In: WEBER, E. (Hrsg.) *Statist. Ausw. Biomed. Daten Teil IV*, Deutsches Krebsforschungsinstitut, Heidelberg.
- HEISELBETZ, C. (1986): PLANUNG: Experimental Planning Facility. *The American Statistician* **40**, 58.
- RASCH, D. (Hrsg.) (1988): *Biometrisches Wörterbuch*. Verlag Harri Deutsch Frankfurt/Main.
- RASCH, D. und HERRENDÖRFER, G. (1982): *Statistische Versuchsplanung*. Dt. Verl. d. Wissensch. Berlin.
- RASCH, D. und HERRENDÖRFER, G. (1986): *Experimental Design. Sample Size Determination and Block Designs*. Reidel Publ. Co. Dordrecht, Boston, Lancaster, Tokyo.

Anschrift des Verfassers: Prof. Dr. Dieter Rasch, Forschungszentrum für Tierproduktion, Dummerstorf-Rostock, DDR-2551 Dummerstorf.

## M. M. Richter

Der Autor greift das populäre Thema der Expertensysteme auf und diskutiert es für seine Bedeutung im Bereich der Statistik. Das ist ein verdienstvolles Unterfangen, weil er für den Laien auf diesem Gebiet (der sich jedoch in der medizinischen Statistik auskennen soll) viele recht aufhellende Bemerkungen macht und vor allem mit einigen weitverbreiteten mißlichen Vorstellungen aufräumt.

Einige zusätzliche, insgesamt die Tendenz des Artikels verstärkende Bemerkungen erschienen mir jedoch angebracht.

Zunächst erscheint es mir unglücklich, den Begriff des Expertensystems definieren zu wollen. Der Verfasser nimmt die strengen Begriffsbildungen später ja auch selber etwas zurück. Ein Begriff wie »Expertensystem« ist informal und entstammt der Umgangssprache. Er wird nicht formal eingeführt, sondern durch zentrale, ihm zukommende Eigenschaften mehr oder weniger näher beschrieben. Vor allem ist ein Expertensystem ein Computerprogramm, das allerdings nicht in einer der herkömmlichen prozeduralen oder algorithmischen Sprachen geschrieben ist. Die verwendeten Sprachen zeichnen sich vielmehr durch ihren deklarativen Charakter aus, was heißt, daß man gewisse Dinge, die man »weiß«, hinschreibt, aber die daraus zu ziehenden Konsequenzen nicht ausprogrammiert, sondern sie dem Inferenzsystem überläßt.

Dabei ist vielleicht nicht ganz klar herausgekommen, daß die Trennung zwischen Daten und Wissen im System selbst keine inhaltliche ist. Wissen spielt sich auf der kognitiven, menschlichen Ebene ab; sobald man es formal repräsentiert, geschieht dies mittels Daten. Wir können jedoch Daten unterscheiden, die fallabhängig sind, und sie einer Datenbank zuordnen, sowie fallunabhängige Daten, deren Aufbewahrungsort man eine Wissensbank nennen könnte.

Von den deklarativen Beschreibungsmethoden diskutiert der Autor vor allem die Regelsprachen. Die Beispiele beziehen sich jedoch auf den etwas vereinfachten, variablenfreien Fall. Etwas zu kurz kommen dabei die heute weitverbreiteten stärkeren Ausdrucksmittel, etwa die Frames. Es soll hier vor allem, angemerkt werden, daß eine Ausgeglichenheit zwischen der Anwendungssituation und der Ausdrucksstärke der verwendeten Sprache zu bestehen hat. Genau wie es für die klassische FORTRAN-Programmierung Probleme gibt, die durch sie völlig befriedigend gelöst werden und die man durch keine Regelsprache verbessern kann, so gibt es entsprechende Situationen auch für die Regelsprachen selber. Andererseits kann aber die Bewältigung mancher komplexer Fragen eben gerade nicht durch solche schwachen Ausdrucksmöglichkeiten befriedigend erledigt werden. In solchen Situationen ein vernünftiges Urteil abzugeben, ist gerade die Aufgabe eines Expertensystem-Designers.

Es gibt in allen Anwendungsbereichen Aufgaben, die man mit einer gewissen Form von Expertensystemen vernünftig behandeln kann, und andere, für die das nicht möglich ist. Genauso gibt es eben auch in der Statistik einfachere und schwierigere Aufgaben. Allgemeine Komplexitätsüberlegungen, weshalb es beispielsweise statistische Expertensysteme gar nicht geben könne, sind hier wenig hilfreich. So wird auch die diskutierte These, daß ein statistisches Expertensystem alle wichtigen Funktionen eines statistischen Experten ausführen kann (und daß es sie deshalb nicht geben kann), von Informatikern sicherlich nicht geteilt. Ein Expertensystem soll bestimmte Aufgaben vernünftig lösen können, und danach wird es rein pragmatisch beurteilt. Charakteristisch ist hier, daß die zu bearbeitenden Aufgaben in der Regel doch einfacher sind, als es ihre erste theoretische Formulierung vermuten läßt. Diese Vereinfachungen drücken sich im sog. »random knowledge« aus, das man zufällig über die auftretenden Situationen hat. Da man es nicht vorhersehen kann, läßt es sich auch nicht a priori ausnutzen. Deklarativ läßt es sich sehr wohl hinschreiben, und die Inferenzkomponente kann daraus ihren Nutzen ziehen.

Abschließend soll noch darauf hingewiesen werden, daß die Methode, Unsicherheit in Form von Faktoren den Regeln beizugeben, viel zu simpel erscheint. Vermutlich müssen hier grundlegend neue Mechanismen entwickelt werden, wodurch sich ein reiches Forschungsgebiet für diejenigen ergeben wird, die sowohl auf dem Gebiet der logischen Wissensrepräsentation als auch auf dem der Statistik ausgewiesene Experten sind.

## B. Streitberg

Herr HAUX hat in seiner Arbeit einen grundlegenden Überblicksartikel vorgelegt zu einem wichtigen und brisanten Thema in der Schnittstelle von Statistik und Informatik. Für diese Leistung möchte ich ihm, gewiß nicht als einziger, meine Anerkennung aussprechen. Herrn HAUX ist es m. E. gelungen, den »state of the art« in gekonnter Weise darzustellen, wobei er auch manches Problematische an diesem Zustand anspricht. Man vergleiche etwa seinen Abschnitt 5, dem ich in vielem völlig zustimmen kann. Freilich sind nach einem Diktum von Voltaire die wertvollsten Arbeiten diejenigen, welche am stärksten zur Kritik anregen. Daher im folgenden einige (kritische) Anmerkungen zu vier Punkten, insbesondere solchen, in denen ich direkt angesprochen wurde:

- (1) Zur Rolle der Komplexitätstheorie
- (2) Zur Inferenz unter Unsicherheit in Expertensystemen
- (3) Konkrete »Expertensysteme« in der Statistik
- (4) Mögliche Probleme für einen Turing-Test stat. Expertensysteme.

### 1. Zur Rolle der Komplexitätstheorie

Herr HAUX hält komplexitätstheoretische Argumente für nicht sonderlich relevant: »Inferenzmechanismen . . . sind relativ unproblematisch. . . liegen größtenteils in . . . P oder NP«. Zunächst ist mir das »oder NP« etwas zu nonchalant – praktisch heißt NP (polynomial zeitbeschränkt auf nichtdeterministischen Automaten) ja, daß ein Problem heute nur in exponentieller Zeit lösbar, also für genügend große Inputs *nicht mehr lösbar* ist. Schließlich sind Computer nur deterministische und keineswegs nichtdeterministische, unbeschränkt vermehrbare Automaten.

Ich würde Herrn HAUX zustimmen, wenn er sagen würde, daß für jede fest vorgegebene Wissensbasis die Komplexitätstheorie ungefähr ebenso relevant ist wie die asymptotische Statistik für die Lösung gegebener finiter Probleme. Die Komplexität ( $n$ ) der Wissensbasis ist dann fest und man kann Algorithmen für Inferenzen in dieser festen Wissensbasis suchen, deren Zeitkomplexität ( $n$ ) akzeptabel ist. Im Regelfall werden diese Algorithmen allerdings nicht mit dem Inferenz-Verfahren einer gegebenen »shell« identisch sein, sondern ganz konkret auf die vorhandene Wissensbasis zugeschnitten sein. HAUX gibt selbst das Beispiel einer Umformung in Entscheidungstabellen, die übrigens m. E. mehr leisten als nur Inferenzen ohne Regelverkettung (denn eine »Entscheidung« kann auch der Rücksprung in die »Inputspalte« sein).

Nun sind Wissensbasen (für genügend schwierige Probleme) im Regelfall nicht starr – es können Fakten oder Regeln hinzugefügt oder beseitigt werden. Will man variable Probleme lösen, ist eine gewisse Veränderbarkeit der Wissensbasis sicher notwendig. In einem PROLOG-ähnlichen Programm heißt das, daß »assert« und »retract« zugelassen sind. Selbst wenn man »assert« so einschränkt, daß nur eine maximale Komplexität des Programms erlaubt wird, landet man bei einem Halteproblem, nämlich dem Problem festzustellen, ob bei beliebigen Anfragen an das System immer noch korrekte und vollständige Inferenzen in der vorgegebenen »akzeptablen« Zeit möglich sind. Ohne starke Restriktionen ist dies ziemlich hoffnungslos, da eine einzige hinzugefügte Klausel das Zeitverhalten eines Inferenzmechanismus katastrophal verändern kann. Für Statistiker: Regelbasierte Inferenz ist »nichtrobust«, der »Break-down-Punkt« ist 0. Natürlich ist FORTRAN auch nichtrobust, aber wer würde den ausführbaren Code eines FORTRAN-Programms zur Laufzeit in Abhängigkeit von Input-Statements beliebig ändern wollen?

In [114] habe ich darauf verzichtet, eine Konsequenz aus den komplexitätstheoretischen Argumenten zu ziehen. Das möchte ich hier nachholen. Entweder ein regelbasiertes System löst nur relativ einfache Probleme, dann kann man sicherstellen, daß es diese auch logisch

korrekt löst. Oder aber man möchte, daß das System auch schwierigere Probleme lösen kann (m. E. sind die Probleme des Biometers schwierig), dann muß man auf »korrekte und vollständige« Inferenz verzichten, also in Kauf nehmen, daß das System unter Umständen *nachweisbar logisch falsche Entscheidungen* trifft. TURING selbst hat das sehr prägnant ausgedrückt: Entweder ist eine Maschine »intelligent« oder »unfehlbar« (*infallible*). Niemals jedoch beides.

Die Komplexitätstheorie ist wichtig, weil sie genau diese Konsequenz aufzeigt. Wenn wir Expertensysteme *wollen*, die zumindest partiell so intelligent sind wie die Experten, die sie modellieren oder ersetzen sollen, dann müssen wir auch akzeptieren, daß diese Systeme notwendig Fehler produzieren werden. Was das »Wollen« betrifft, habe ich selbst gemischte Gefühle. Personalentscheidungen, Arzneimittelzulassungen, Therapieempfehlungen würde ich ungern solchen Systemen überlassen (bei Fehlentscheidungen eines Menschen gibt es immerhin einen Verantwortlichen), bei der Auswahl eines »Tototips« wäre ich gewiß toleranter.

Wenn es richtig ist, daß die Konstruktion »intelligenter« Programme notwendig falsche Programme produziert, wozu dann regelbasierte Systeme? Also warum soll man erst mühsam viele Regeln in einer logischen Sprache aufstellen, wenn das System dann ohnehin in unvorhersagbarer Weise falsche Schlüsse daraus zieht? Zudem ist m. E. ein »regelbasierter« Statistiker höchstens zum Bestehen der ersten Statistik-Klausur geeignet. Die Lektüre von IMMICHs »Medizinischer Statistik« oder von LEHMANS »Testtheorie« würde er wohl genausowenig schaffen wie die Lösung des ersten praktischen Problems, welches man ihm stellt. Oder glaubt jemand ernsthaft daran, daß etwa Herr IMMICH ein besonders umfangreiches regelbasiertes System darstellt?

Nun gibt es eine Alternative, die von vornherein auf Logik und Richtigkeit verzichtet und diese durch stochastische Approximationsverfahren ersetzt. Dies sind *neuronale Netze*, also hochvernetzte Systeme parallel arbeitender einfacher Einheiten. Die Statistik kommt hier dadurch hinein, daß zum einen die Einheiten selbst stochastischer Natur sein können (sog. »Boltzmann Orakel«) und daß zum zweiten die Systeme stochastisch trainiert werden (durch zufällige Präsentation von Stimuli aus einer vorgegebenen Menge). Unter bestimmten Voraussetzungen »konvergieren« diese Systeme gegen etwas »Vernünftiges«, d. h. gegen etwas, was wir auch als Statistiker akzeptieren können (eine »minimal distance partition«, eine Fishersche Diskriminanzlösung oder den Kleinstquadrateschätzer). In der Praxis konvergieren neuronale Netze jedoch i. a. nur bezüglich bestimmter Entscheidungen, andere bleiben offen (man kann zeigen, daß hier oft fraktale Attraktoren auftreten, deren Hausdorff-Besicovitch-Dimension echt größer als die topologische Dimension ist, d. h. gewisse Projektionen der Netzzustände bewegen sich auf Mengen wie dem Cantorsche Diskontinuum oder dem Sierpinski-Schwamm). Man kann neuronale Netze auch mit Hilfe von Fakten und Regeln (geeignet als Stimuli bzw. gewünschte Stimulus-Output-Relationen codiert) trainieren. Wo immer man das versucht hat, ergaben sich angebliche Resultate, die mit viel geringerem Aufwand mindestens so brauchbar waren wie regelbasierte Systeme. Meine konkrete Folgerung wäre, ob wir nicht versuchen sollten, die Probleme stat. Expertensysteme mit Hilfe der Statistik selbst, also geeigneter neuronaler Netze, in den Griff zu bekommen.

## 2. Inferenz unter Unsicherheit

HAUX zitiert hier eine (vielleicht etwas bajuwarische) Verbalinjurie von mir: »Simply pseudoscientific mumbo jumbo«. Diese bezog sich jedoch nur auf diejenigen Methoden, die in den meistpropagierten Systemen verwendet werden. Leider sind die wissenschaftlich äußerst wertvollen Untersuchungen von HAJEK, DEMPSTER und anderen zu diesem Thema noch kaum praktisch eingesetzt worden. Diese Systeme benutzen vielmehr, nach einem Satz

von HAJEK, i. a. mehr oder minder irrational begründete Umformungen von »Idiot Bayes« (also Bayes-Verfahren mit Unterstellung der Unabhängigkeit aller Parameter). Das Problem ist naheliegenderweise sehr schwierig: Man möchte subjektive Wahrscheinlichkeiten von Fakten berechnen, gegeben die Wahrscheinlichkeit anderer Fakten und die »Wahrscheinlichkeit« der Regeln, wobei über die gemeinsame Verteilung nichts oder nur wenig bekannt ist. Eine konservative Methode könnte hier Intervallgrenzen für die subjektiven Wahrscheinlichkeiten propagieren. Dies geschieht z. B. im System INFERNO, wobei der Name recht passend gewählt wurde, da man sehr schnell die Intervallgrenzen  $[0,1]$  erhält.

### 3. Konkrete statistische Expertensysteme

HAUX widerlegt mich recht summarisch: es gibt stat. Expertensysteme (etwa GLIMPSE). Also ist der Satz, daß es keine geben kann, falsch. Ich muß allerdings gestehen, daß mich die konkreten Beispiele, die er bringt, nicht überzeugen (der Begriff »Experte« scheint hier doch sehr abgemagert). Man vergleiche z. B. GLIMPSE bei der Auswahl der Variablen  $X_1, X_2, X_3 \dots$  in einem generalisierten linearen Modell (Abbildung 4). Man beachte übrigens, daß Probleme der Variablenselektion einer math. Analyse durchaus zugänglich sind (vgl. etwa schwache Konsistenz von AIC oder starke Konsistenz von BIC etc.).

Der aktuelle Zustand von Expertensystemen in der Statistik scheint mir durchaus vergleichbar der Situation von Expertensystemen zur Kindererziehung. Ein derartiges System wurde in einem der letzten BYTE-Hefte besprochen. Man sagt dem Programm, daß der Säugling schreit. Es fragt, ob er schon länger als 15 min schreit. Man muß bejahen (er schreit schon die ganze Nacht). Das System fragt, ob der Säugling immer noch schreit, wenn man ihn aus dem Bettchen nimmt. Er tut es. Das System denkt nun einige Zeit nach und dann kommt die Entscheidung in Form eines bedrohlich rot blinkenden Bildschirms: »Dringend! Dringend! Nehmen Sie sofort Kontakt mit einem Arzt auf. Sonst könnte sich eine furchtbare Tragödie ereignen!« Glücklicherweise ist der Säugling von der roten Farbe so fasziniert, daß er das Schreien vergißt und bald friedlich einschläft.

### 4. Ein TURING-Test für statistische Expertensysteme

Dankenswerterweise hat Herr HAUX meinen Vorschlag eines »TURING-Tests« für statistische Expertensysteme aufgegriffen und m. E. konkret verbessert. Auf einer der nächsten Reisenburg-Sitzungen soll nun der Kampf »Mensch gegen Maschine« stattfinden. Einer der Herausforderer ist das Programm »AUTOBOX«, das automatische BOX-JENKINS-Analysen von Zeitreihen liefert. Gesucht werden menschliche Gegner, die sich dem Schiedsrichter (Prof. Dr. P. NAEVE, Universität Bielefeld) namentlich melden sollten. Sofern ein Mensch gegen das Programm gewinnt, erhält er als Preis eine Gratis-Kopie des Programms. Ich hatte dem Schiedsrichter begeistert eine EEG-Reihe annonciert, die ab einem bestimmten Moment vom Effekt eines Psychopharmakons beeinflusst wird. Auf die Länge der Zeitreihe befragt, mußte ich allerdings beschämt eingestehen, daß mit  $n > 100\,000$  zu rechnen sei. Das das Expertensystem nur Reihen mit  $n < 300$  behandelt, muß ich wohl leider auf altmodische Programmiermethoden zurückgreifen.

Ein Expertensystem in der Biometrie sollte Probleme behandeln können, die ein Experte in der Biometrie normalerweise behandelt (vgl. die Definition von HAUX in 3.1/2.). Ein solches Problem möchte ich abschließend stellen (mehr oder minder willkürlich aus den Problemen auf meinem Schreibtisch herausgegriffen). Daten sind ein  $(93, 28, 64)$ -Array: für 93 Probanden wurden an 28 Orten auf dem Schädel EEG-Ableitungen erfaßt und spektral analysiert, gegeben sind die mittleren Spektralampplituden in 64 Frequenzbändern (0–0,4 Hz, 0,4–0,8 Hz. etc.). Das Ziel der Untersuchung ist rein explorativ: Da hinter dem Array

ein ungeheurer meßtechnischer und rechentechnischer Aufwand steckt, interessiert man sich für eine Datenreduktion – reichen unter u. U. weniger Elektroden aus und ist eine gröbere Zerlegung der Frequenzachse statthaft? Interessant an dem Datensatz ist ein Defekt, den erst eine eingehende statistische Analyse zeigen konnte: gewisse Parameter wiesen sehr große Ähnlichkeit von »Probanden« mit aufeinanderfolgenden Nummern auf (etwa Nr. 40 und 41 etc.) Dies legte die Vermutung nahe, daß es sich gar nicht um unabhängige Probanden handelt, sondern Wiederholungsmessungen eingeschlossen wurden. Der Anwender mußte zu seinem Entsetzen diese Vermutung bestätigen.

Ein Expertensystem, das den beschriebenen Defekt von sich aus findet (natürlich ohne vorher über diese Möglichkeit informiert worden zu sein) und noch irgend etwas nicht allzu Irrationales für die Fragestellung tut, würde ich als eine echte Widerlegung meiner These von der Nichtexistenz statistischer Expertensysteme akzeptieren. Meines Erachtens sind wir jedoch noch weit von derartigen Systemen entfernt.

Anschrift des Autors: Prof. Dr. B. Streitberg, Alt-Moabit 90, 1000 Berlin 21

## K. M. Wittkowski

HAUX [4] hat zu der Diskussion über »statistische Expertensysteme« einen wesentlichen Beitrag geleistet, indem er in seiner Definition nicht festlegt, ob sie Entscheidungsprozesse von Menschen »simulieren« bzw. »nachbilden« oder ob sie Probleme auf andere Weise lösen, welche Programmiersprache und Hardware verwendet wird und ob die Mensch-Computer-Schnittstelle »natürlich-sprachlich« ist. Eine Definition über die vom System lösbaren Probleme ist aus vielen Gründen vorzuziehen: Niemand würde von einem Flugzeug erwarten, daß es die Flugtechnik von Vögeln »simuliert«. Die Implementierung von PUFF in BASIC hat gezeigt, daß es »Expertensysteme« mit gleicher Funktionalität in verschiedenen Programmiersprachen geben kann [8]. Experten eines Fachgebietes verwenden oft eine wenig »natürliche« Sprache zur Kommunikation. Die von HAUX gewählte Begriffsbestimmung ist jedoch in einer Hinsicht etwas zu weit gefaßt: Wenn ein System ausschließlich arithmetische oder mengentheoretische Operatoren auf Daten anwendet, dann sollte man es auch dann nicht als »Expertensystem« bezeichnen, wenn diese Funktionen bisher von Experten »manuell« durchgeführt werden mußten. Ein Expertensystem unterscheidet sich dadurch von einem Auswertungssystem, daß es nicht nur Daten, sondern auch Semantik, d. h. Wissen über die Daten, verarbeitet.

### Definition

*Ein Programm ist ein »Statistisches Expertensystem«, wenn es durch die Anwendung von Wissen über Daten spezifische Expertentätigkeit in der Statistik oder in einem Teilgebiet hiervon übernehmen kann.*

In der Vergangenheit lag das überwiegende Interesse an »Expertensystemen« in der Entwicklung von Grundlagen und Tools, d. h. auf dem Gebiet der Informatik. Es ist deshalb nicht verwunderlich, daß bisher nur wenige dieser Systeme erfolgreich zur Lösung relevanter Probleme in der Praxis eingesetzt wurden. Wenn man Expertensysteme nunmehr anhand der von ihnen lösbaren Probleme definiert, dann ergeben sich daraus unmittelbar die von HAUX formulierten Anforderungen einerseits an die Systeme (das Problem

bestimmt die Auswahl eines Shells, und nicht umgekehrt) sowie andererseits an Qualität des gespeicherten Wissens (eine unkritische Übernahme von »Kochrezepten« führt mit Sicherheit eher zu mehr Fehlinterpretationen als zu einer Reduktion von Fehlerquellen). Nur bei hohen Anforderungen an das implementierte Wissen lassen sich solche Fehlentwicklungen vermeiden. Der Grundsatz »Garbage in – garbage out« wird auch durch die Anwendung von Expertensystemen (»rapid prototyping«) nicht aufgehoben. Zwei zentrale Aussagen von HAUX sollten deshalb als Postulate hervorgehoben werden:

#### Postulat 1

*Für die Wissensbasis eines statistischen Expertensystems müssen Experten auf den jeweiligen Teilgebieten der Statistik verantwortlich zeichnen.*

#### Postulat 2

*Das implementierte Wissen muß so offengelegt werden, daß es überprüft und mit den verantwortlichen Experten diskutiert werden kann.*

Nur wenn diese Bedingungen erfüllt sind, kann eine inhaltliche Diskussion über die Qualität des Wissens anhand der expliziten Repräsentation geführt werden. Der von HAUX (nach STREITBERG) vorgeschlagene TURING-Test könnte dagegen unter Umständen in die Irre führen, solange es unter »Experten« unterschiedliche »Ansichten« gibt. Die von HAUX angeführten Überlegungen zum Vergleich von (unverbundenem) t-Test und WILCOXON-Test sind z. B. vielen »Experten« offensichtlich unbekannt, obwohl sie bereits seit langem diskutiert werden [6, 2, 10]. In einer solchen Situation könnte ein Expertensystem anhand des implementierten Wissens eines auf seinem Spezialgebiet höher qualifizierten Wissensverwalters von den »Experten« als abweichend erkannt werden, weil die Entscheidungen des Systems richtig und die Entscheidungen der menschlichen »Experten« falsch sind. Da Wissenschaft ihrem Wesen nach undemokratisch ist, darf die mehrheitliche Ablehnung des Lösungsvorschlages eines Expertensystems durch ein Gremium von »Experten« nicht automatisch zu einem negativen Urteil führen (vgl. z. B. die Diskussion über Empfehlungen zur Anwendung von bedingten und unbedingten Statistiken zum Test von Tendenzhypothesen [5, 11–13]). Die Diskussion über Expertensysteme sollte gegebenenfalls zum Anlaß genommen werden, »allgemein anerkannte« Kochrezepte zu überprüfen und gegebenenfalls zu korrigieren.

Die aktuelle Entwicklung belegt, wie notwendig solche präzisen Begriffsbestimmungen und Anforderungen sind. Es ist z. B. noch nicht klar, ob KENS und CADEMO nach der obigen Definition Expertensysteme sind. »KENS could be described as halfway between a question answering system and an information retrieval system« [3]. Sowohl die Anpassung einer Wachstumsfunktion an Daten anhand der Restvarianz als auch die Bestimmung »optimaler« Stützstellen [7] sind zwar aufwendig, erfordert jedoch kein Wissen über die Daten. REX und GLIMPSE verwenden ebenfalls kein Wissen bzw. nur sehr wenig Wissen über Daten. Die »Verfügbarkeit« von REX ist inzwischen weiter eingeschränkt, da GALE die Arbeiten an REX und STUDENT eingestellt hat. Die Idee, ein System von seinem Benutzer »lernen« zu lassen, habe sich als nicht realisierbar herausgestellt [persönliche Mitteilung, 2nd International Workshop on AI and Statistics, Fort Lauderdale, FL, 1989. Die Tagungsbeiträge sollen in den *Annals of Mathematics and Artificial Intelligence* sowie Anfang 1990 als Buch erscheinen]. Da die meisten Anwender von Auswertungssystemen »Fachpersonen« ohne Expertenwissen in Statistik sind, ließe sich ein solcher Ansatz auch kaum mit dem obigen Postulat 1 vereinbaren. Ähnliche Probleme sind zu erwarten, wenn Benutzer P-STAT tatsächlich als »expert system shell« [BUHLER, persönliche Mitteilung, 4.

Konferenz über die wissenschaftliche Anwendung von Statistik-Software, Heidelberg, FRG, 1988] nutzen. Obwohl die Diskussion über angemessene Problemlösestrategien und Wissensbasen noch lange nicht abgeschlossen ist, werden seit neuestem mindestens zwei weitere »Expertensysteme« kommerziell angeboten: TURBO SPRING-STAT [1] sowie ESS [8]. Der Beitrag von HAUX kann hier wertvolle Anhaltspunkte zur Evaluation geben. Gefördert mit Mitteln des NIH (International Research Fellowship F05TWO4087)

## Literatur

1. BROWN, B. (1989): Statistical analysis – Turbo Spring-Stat, PC Magazine **8**, 159–162.
2. ELASHOFF, J. D., R. M. ELASHOFF (1977): Effects of errors in statistical assumptions. Intern. Enycl. Stat. **1**, 229–250.
3. HAND, D. (1987): A statistical knowledge enhancement system. J. Royal Stat. Soc. A **150**, 334–345.
4. HAUX, R. (1989): Statistical expert systems. Biometrie und Informatik in Medizin und Biologie (dieses Heft).
5. HAUX, R., R. HILGERS, A. HÖRMAN, K. H. JÖCKEL, W. LEHMACHER (1988): Recommendations on software for non-parametric statistical methods. Statistical Software Newsletter **14**, 27.
6. LEHMANN (1975): Nonparametrics – statistical methods based on ranks. San Francisco, CA: Holden-Day.
7. RASCH, D. (1988): CADEMO – ein Expertensystem für die Versuchsplanung und die Wahl statistischer Auswertemodelle, VEB Elektro-Consult: Berlin, DDR.
8. THOMSON, E. (1989): Statistical system does the thinking for users. Digital review, Feb 6, 31–34.
9. WATERMAN, D. A. (1986): A guide to expert systems. Reading, Mass: Addison-Wesley.
10. WITKOWSKI, K. M. (1986): An expert system for testing statistical hypotheses: strategies for knowledge engineering. In: BOARDMAN, T. I. (ed.) Computer science and statistics, Washington, DC: ASA, 438–443.
11. – (1988): Friedman-type statistics and consistent multiple comparisons for unbalanced designs. J. Am. Stat. Assoc. **83**, 1163–1170.
12. – (1989): An asymptotic UMP sign test for discretized data. The Statistician (in press).
13. – (1989): A note on recommendations regarding the handling of ties in software for nonparametric statistical methods. Statistical Software Newsletter **15** (in press).

Anschrift des Verfassers: Inst. für Med. Biometrie der Eberhard-Karls-Universität, D-7400 Tübingen und Dept. Comp. Sci., University of Pittsburgh, PA 15260

## Abschließende Stellungnahme des Verfassers

Sämtliche Kommentare enthalten meines Erachtens wichtige und interessante Ergänzungen zu den – wie von mir in Abschnitt 1 erwähnt – einführenden und keinesfalls den Anspruch auf Vollständigkeit erhebenden Ausführungen. Bevor ich auf einige Aspekte der Kommentare im einzelnen eingehen möchte, sei hier noch einmal betont, daß meines Erachtens Produktionsregeln und ähnlich primitive Wissensrepräsentationsformen (mit ihren Inferenzmechanismen) im allgemeinen nicht geeignet sind für die formale Repräsentation statistischen Wissens. Nach meiner Ansicht wird auch in mehreren Kommentaren viel zu sehr auf die »Fakten« und »Regeln« im speziellen und nicht auf Wissensrepräsentationsformen im allgemeinen eingegangen.

Wie W. G. DORDA sehe auch ich die mögliche Gefahr der »Suggestion falscher Sicherheit« durch Statistische Expertensysteme. Diese läßt sich nur vermeiden, wenn das Systemverhalten transparent und reproduzierbar ist, und wenn Wissen explizit repräsentiert und nicht im Programmcode versteckt wird. Jeder Benutzer eines Statistischen Expertensystems sollte beispielsweise die Möglichkeit haben, sich den kompletten Inhalt der Wissensbank ausdrucken lassen zu können. Jeder Autor einer Arbeit über ein konkretes Statistisches Expertensystem sollte, in Analogie dazu, sein repräsentiertes Wissen mit veröffentlichen oder zumindest verfügbar machen. Ein Verbergen von Wissen, oft verbunden mit persönlichen Besitzansprüchen, halte ich für einen Rückschritt wissenschaftlichen Arbeitens.

F. EICKER und zahlreiche andere Kommentatoren schlagen vor, den Begriff Statistisches Expertensystem »tunlichst zu unterdrücken«. Auch dies möchte ich begrüßen. Nach meinen Erfahrungen auf dem Gebiet der wissensbasierten Diagnose- und Therapieunterstützung in der Medizin läßt sich dies dann erreichen, wenn an konkreten Systemen für konkrete Probleme gearbeitet wird. Allerdings läßt sich der Begriff auch dann nicht immer ganz umgehen. Ich möchte um Nachsicht bitten, wenn ich ihn deshalb hier dennoch, notgedrungen, weiter verwende.

K. A. FRÖSCHL bemerkt, daß der Begriff »spezifische Expertentätigkeit . . .« »selbst recht unspezifisch« ist. Um diesen auch von mir in Abschnitt 5.1 angedeuteten Sachverhalt noch weiter zu belegen, erlaube ich mir, folgendes zu ergänzen: In der 17. Auflage der BROCKHAUS-Enzyklopädie steht, daß ein »Experte« ein Sachverständiger ist. Schlägt man nach, was man unter einem »Sachverständigen« verstehen soll, so heißt es dann, daß dieser ein Fachmann ist. Unter dem Begriff »Fachmann« findet sich dann der Hinweis, man möge die Begriffsbestimmung unter dem Stichwort »Experte« nachschlagen.

K.-H. JÖCKEL und R. OSTERMANN schlagen vor, nicht nur theoretische Überlegungen zur Adäquatheit zu machen, sondern empirische Untersuchungen über Statistische Expertensysteme durchzuführen. Auch dies scheint mir wichtig und nunmehr auch möglich zu sein, zumal solche systematisch geplanten, durchgeführten und ausgewerteten Untersuchungen einem Statistiker besonders liegen sollten.

Zu Abschnitt 5 in dem Kommentar von H.-J. LENZ sei noch ergänzt, daß mittlerweile schon mehr als die zwei von mir erwähnten Statistischen Expertensysteme kommerziell verfügbar sind (vgl. den Kommentar von D. RASCH). Es werden wohl noch weitere hinzukommen – allerdings dürfte mit zunehmender Zahl verfügbarer Statistischer Expertensysteme auch die Erwartung, was solche Systeme leisten sollen, weiter abnehmen.

I. W. MOLENAAR fragt, ob es richtig ist, daß PANOS, im Gegensatz zum »liberalen« GLIMPSE, »autoritär« ist. Dies mit dem Hintergrund, daß PANOS für Fachpersonen sein soll, und nicht für Laien. K. M. WITKOWSKI ist in der Tat der Ansicht, daß auch Statistische Expertensysteme für eine statistische Fachperson »autoritäre Verhaltensweisen« – zur Qualitätsverbesserung statistischer Datenanalysen – haben sollte. Diese Ansicht wird jedoch sehr konträr diskutiert. Professor MOLENAAR schlägt vor, daß sich wissenschaftliche Gremien aktiv an der Bewertung der Qualität Statistischer Expertensysteme beteiligen. Diese Aktivität scheint mir außerordentlich wichtig zu sein, gleichzeitig bin ich allerdings leider auch ratlos, wie diese Bewertung konkret geleistet werden könnte.

P. NAEVE kommentiert meinen tapferen Versuch, sich mit der keinesfalls steinharten Materie der Statistischen Expertensysteme zu befassen, u. a. mit einem Gedicht von GOETHE.

Er kritisiert meine »schwammige Begrifflichkeit« bei der Begriffsbestimmung von Expertensystemen. Um nicht ganz so schwammig zu bleiben, sei hier noch folgende Ergänzung hinzugefügt: Ich wollte – wenn überhaupt – den Begriff Expertensysteme nicht anhand irgendwelcher Architektur Aspekte (»ein Statistisches Expertensystem hat eine Wissensbank«) oder gar Sprachaspekte (»ein Statistisches Expertensystem ist in LISP oder in PROLOG geschrieben«), sondern mittels seiner Funktionsweise (real-) definieren. Es ist meines Erachtens beispielsweise unsinnig zu sagen, daß jedes regelbasierte Programm ein

Expertensystem sei. Dennoch gibt es auch solche Definitionsversuche, die damit auch ein einfaches (in Produktionsregeln geschriebenes) Sortierprogramm zum Expertensystem erheben. Deshalb meine Bedingung: ein Expertensystem sollte *spezifische* Expertentätigkeiten übernehmen können. Dies sind Tätigkeiten, die eben nur von Experten übernommen werden können. Lassen sich diese aber auf eine Maschine abbilden, dann können diese spezifischen Expertentätigkeiten von Experten und außerdem von solchen Maschinen übernommen werden. Den Zirkelschluß, daß damit jede Maschine automatisch zum Expertensystem wird, mag aus meinem Definitionsversuch interpretiert werden können. Er war aber nicht beabsichtigt.

Nun hat Professor NAEVE einen ebenso tapferen Gegenvorschlag für eine Begriffsbestimmung gemacht, indem er die von D. J. HAND vorgeschlagene Definition als akzeptabel beschreibt. Dieser Vorschlag enthält jedoch, neben der Vermengung von notwendigen und gewünschten Eigenschaften Statistischer Expertensysteme, die explizite Einschränkung auf regelbasierte Programme. Diese Einschränkung wird sicherlich nicht nur von mir als falsch und unnötig erachtet. Verweisen möchte ich in diesem Zusammenhang auf den Kommentar von M. M. RICHTER. Um diesen Disput hier zu beenden, sei mir erlaubt – mit Bezug auf den zweiten Begriffsvorschlag – PETER NAEVE mit der von meinem schwäbischen Landsmann BERTOLT BRECHT verfaßten zweiten Strophe des Liedes von der Unzulänglichkeit menschlichen Strebens aus der Dreigroschenoper zu antworten:

Ja, mach nur einem Plan  
Sei nur ein großes Licht!  
Und mach dann noch 'nen zweiten Plan  
Gehn tun sie beide nicht

(die erste Strophe möchte ich im Zusammenhang mit Statistischen Expertensystemen lieber ganz verschweigen).

Zu D. RASCH möchte ich ergänzen, daß PANOS nicht für die Versuchsauswertung, sondern schon für die Versuchsplanung konzipiert wurde.

M. M. RICHTER schreibt zu Recht, daß sich die Trennung von Daten und Wissen nicht immer durchführen läßt und daß es keine klar definierte inhaltliche Trennung gibt.

B. STREITBERG erscheint meine Aussage zweifelhaft, daß Statistische Expertensysteme, die in die Komplexitätsklasse NP gehören, unproblematisch sind. Sicherlich waren für mich hier die eigenen Erfahrungen mit jeweils fest vorgegebenen Wissensbasen (die natürlich von Zeit zu Zeit erweitert oder modifiziert werden) ausschlaggebend. Und insofern herrscht Konsens zwischen uns. Abschnitt 3 des Kommentars, zusammen mit Abschnitt 4, zweiter und dritter Absatz, zeigen nochmals die unterschiedliche Sichtweise dessen, was ein Statistisches Expertensystem leisten soll. Auch ich kann mir übrigens nicht vorstellen, daß ein statistischer Experte mit einem regelbasierten System gleichgesetzt werden könnte. Andererseits möchte ich mit M. M. RICHTER argumentieren, daß es ja darum nicht geht, sondern vielmehr darum, daß ein Statistisches Expertensystem ja »nur« bestimmte Aufgaben vernünftig lösen können« soll.

K. M. WITKOWSKIS Postulat 2 sollte auch bei der Herausgabe von Zeitschriften Berücksichtigung finden: Arbeiten zu konkreten Statistischen Expertensystemen sollten nur dann zur Veröffentlichung angenommen werden, wenn das formal repräsentierte Wissen offen liegt, sei es, daß es mit veröffentlicht wird, sei es, daß es für jeden Leser erhältlich ist.

## An exact nonparametric randomization test for censored data

R. A. Hilgers and N. Neumann

### Summary

*In this paper an exact two-sample randomization test under a global null hypothesis of interchangeability or equality of distributions for arbitrarily censored data is proposed and developed from a practical example. The tests discussed by GEHAN (1965a, 1965b) for singly-right- and doubly-censored data are just approximate asymptotic versions of our procedure. Thus the variances needed for their appropriate standardization may be computed in a more straightforward and easy way rather than using GEHAN'S somewhat complex formulas.*

### Zusammenfassung

*Es wird ein exakter Zwei-Stichproben-Randomisierungstest unter der globalen Nullhypothese der Austauschbarkeit bzw. der Gleichheit der Verteilungen für beliebig zensierte Beobachtungen vorgeschlagen und an einem praktischen Beispiel entwickelt. Die von GEHAN (1965a, 1965b) diskutierten Verfahren für rechts- und doppelt-zensierte Beobachtungen sind dabei asymptotisch-approximative Versionen in speziellen Zusammenhängen für unseren Test. Daher können die zu Standardisierungen dabei erforderlichen Varianzen auch in einer einfachen und natürlichen Weise aus dem Randomisierungsansatz berechnet werden, anstatt die für die Praxis doch etwas unhandlichen Formeln von GEHAN zu verwenden.*

### Key words

*nonparametric two-sample tests, randomization test, exact distribution, GEHAN'S tests, censored observations, interval-valued observations*

### 1. Introduction

In practical statistics censored data play an important role. The best known example for such a type of observation are the survival times where the interesting true life time of an individual under some regimen may be greater than the actually observable time under study. But in practice almost all observations which theoretically are on a continuous scale in fact are censored either due to the limitations of the equipment used or by substantial reasons.

Thus it will be of practical importance to have a test procedure for the comparison of samples where observations may be censored either at one side, left or right, or even are doubly censored, i.e. in other terminology they are interval-valued.

GEHAN (1965a, 1965b) proposed nonparametric procedures in the context of survival times for either singly- or doubly-censored data. But these procedures are only asymptotical results and may not be applied for rather small samples.

In this paper we propose an exact randomization test which is the small sample case of GEHAN'S proposals and which has the further advantage that the variance estimator needed for standardization in an approximate version is much easier to compute than by the formulas provided by Gehan. This is due to the fact that his estimators derived under the null hypothesis of equal distributions coincide with the ordinary permutation variance.

In the next section we shortly recapitulate the well-known procedure for singly-censored data from GEHAN (1965a) and give a further example with censored data from practice in sec. 3. For this we will develop the exact finite procedure in sec. 4 from an intuitive point of view. A mathematical justification is given in sec. 5 before in sec. 6 the procedure is applied to our example.

**2. GEHAN'S Approximate Test for Singly-Censored data**

Since our aim is a conditional procedure, given the observations, we just take a conditional point of view rather than discussing random variables on some measurable space. Let  $x_{ij}$  be real valued observations,  $i = 1, 2, j = 1, 2, \dots, n_i$ , possibly censored at the right, i.e. the value of interest is at least as great as the value actually measured. Then GEHAN'S test statistic is given as the standardized quantity

$$(2.1) \quad T = \frac{H^G}{S}$$

where  $S^2$  is some estimator of variance which will be discussed later on from another point of view and  $H^G$  is the U-statistic

$$(2.2) \quad H^G = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} u(x_{ij_1}, x_{2j_2})$$

with the comparison function

$$(2.3) \quad u(x, y) = \begin{cases} +1 & x > y \\ -1 & x < y \\ 0 & \text{else} \end{cases}$$

Here  $\gg\ll$ ,  $\ll\ll$  denote a binary relation of strict order and its dual, respectively, which in the present case is given by

$$(2.4) \quad x > y \iff \begin{cases} x > y \text{ and } y \text{ uncensored} \\ \text{or} \\ x = y \text{ and } x \text{ censored and } y \text{ uncensored} \end{cases}$$

$$x < y \iff y > x$$

Thus  $u(\cdot, \cdot)$  is to be interpreted as the difference of two indicator functions for strict order and its dual, respectively, cf. HILGERS (1988) for further details.

For what he calls  $\gg$ grouped data $\ll$  Gehan considers a slightly different strict order, namely

$$(2.5) \quad x > y \iff x > y \text{ and } y \text{ uncensored}$$

The difference between both definitions is the simple fact that in (2.4) a censored observation which numerically is equal to an uncensored one is considered to reflect a better survival which in fact is justified if the measurement is rather exact, say days. On the other hand if measurement is rough by itself, say only month, then (2.5) is preferable, since with this definition numerical equality is judged as indifference.

The test statistic  $H^G$  in (2.2) by definition may be represented essentially as a  $n_1 \times n_2$  matrix with entries  $+1, -1, 0$ , respectively, which is summed up over all columns and then over all rows. In the next section we will give an example for censored data other than survival times along with a first demonstration of the computation of the test statistic which will lead us to the solution for the exact test.

### 3. Another Example for Censored Data

To demonstrate a situation where data typically are censored other than survival times we give a small example of two samples of independent observations where the quantity of interest is the titer against some antibody.

Table 1. Measurements for two independent samples of antibody titer (VCA: Viral Capsid Antigen). I: Infectious Mononucleosis. II: Nasopharynx-Carcinoma.

I	II
1: 640	1:2560
1:1280	1:1280
1: 640	1: 320
1: 160	1:2560
<1: 80	1:1280
1: 320	1:5120
	>1:5120

In this example we have left ( $<1:80$ ) as well as right ( $>1:5120$ ) censored observations which may be represented equivalently as intervals  $[0, 80)$  and  $[5120, \infty)$ , respectively, in which the true but unknown exact titer is situated. Furthermore, all other measurements must be regarded as being doubly-censored, e.g. for »1:1280« the only we know is that the true value is at least »1:1280« but less than »1:2560«. Thus this notation is a representation for the interval of dilution  $[1280, 2560)$ . Table 3.2 gives the measurements in interval notation.

Table 2. Measurements for two independent samples of antibody titer in interval notation.

I	II
[160, 320)	[1280, 2560)
[640, 1280)	[ 640, 1280)
[320, 640)	[ 320, 640)
[ 80, 160)	[1280, 2560)
[ 0, 80)	[ 640, 1280)
[160, 320)	[2560, 5120)
	[5120, $\infty$ )

Since all observations are equivalently represented by intervals which are left closed and right open, i.e.  $[x_l^i, x_r^i)$ , the strict order is given as

$$(3.1) \quad x > y \iff x^l \geq y^r$$

Then applying the comparison function (2.3) to this strict order for all pairwise inter-sample comparisons yields the matrix

Table 3. Matrix of the inter-sample comparisons with the function  $u(\cdot, \cdot)$  from (2.3) applied for the strict order (3.1).

i	j	2							$\Sigma$
		1	2	3	4	5	6	7	
1	1	-1	-1	+1	-1	-1	-1	-1	-5
	2	-1	0	+1	-1	0	-1	-1	-3
	3	-1	-1	+1	-1	-1	-1	-1	-5
	4	-1	-1	-1	-1	-1	-1	-1	-7
	5	-1	-1	-1	-1	-1	-1	-1	-7
	6	-1	-1	0	-1	-1	-1	-1	-6

The sum of all entries results in the finite test statistic  $H^G = -33$ . Since sample sizes are as small as 6 and 7 observations, respectively, it does not seem appropriate to use an asymptotic approximation. Nevertheless, in section 6 we will give a simple formula for the correct estimator of variance which has to be used for standardization if samples are such large that the computation of the exact distribution or only the exact tail probabilities might become too time-consuming.

From a theoretical point of view there will be no problem at all to establish an exact randomization test just from definition. We only have to compute the statistic for all possible combination of 6 and 7 observations out of the total number of 13 which under the nullhypothesis of interchangeability or the equality of distributions are equally likely. But,

there will be  $\binom{13}{6} = \binom{13}{7} = 1716$  possible combinations with  $6 \cdot 7 = 42$  pairwise comparisons

each. Thus even for this rather small example there will arise some computational difficulty with respect to time.

#### 4. Reduction of the Problem

To establish a permutation test at a reasonable amount of effort, we reduce the problem by, as is common in mathematics, enlarge the problem. The statistic  $H^G$  is based on the pairwise comparisons from all observations of the first to all of the second sample, the inter-sample comparisons, and then summing up. Let us perform the intra-sample comparisons as well, i.e. all observations of a sample are compared to all of the same sample. For our example we thus get the matrix

Table 4. Intra- and inter-sample comparisons with row sums and their squares for the example of antibody titer.

i	j	1					2							$\Sigma$	$\Sigma^2$	
		1	2	3	4	5	6	1	2	3	4	5	6			7
1	1	0	-1	0	+1	+1	+1	-1	-1	+1	-1	-1	-1	-1	-3	9
	2	+1	0	+1	+1	+1	+1	-1	0	+1	-1	0	-1	-1	+2	4
	3	0	-1	0	+1	+1	+1	-1	-1	+1	-1	-1	-1	-1	-3	9
	4	-1	-1	-1	0	+1	-1	-1	-1	-1	-1	-1	-1	-1	-10	100
	5	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-12	144
	6	-1	-1	-1	+1	+1	0	-1	-1	0	-1	-1	-1	-1	-7	49
2	1	+1	+1	+1	+1	+1	+1	0	+1	+1	0	+1	-1	-1	+7	49
	2	+1	0	+1	+1	+1	+1	-1	0	+1	-1	0	-1	-1	+2	4
	3	-1	-1	-1	+1	+1	0	-1	-1	0	-1	-1	-1	-1	-7	49
	4	+1	+1	+1	+1	+1	+1	0	+1	+1	0	+1	-1	-1	+7	49
	5	+1	0	+1	+1	+1	+1	-1	0	+1	-1	0	-1	-1	+2	4
	6	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	0	0	-1	+10	100
	7	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	0	+12	144

From this we take the vector of all row sums and, for later use, the vector of the squares of the row sums, cf. (5.7). As test statistic  $H^P$  we then will use the sum of the first  $n_1$  row sums, i.e. the sum of all comparisons of observations from the first sample to all observations, irrespective of whether these belong to the first or the second sample.

The resulting  $H^P = -33$  is numerically equal to the value found from the inter-sample comparisons alone. This on the other hand is not surprising since the sum of all intra-sample comparisons must be zero. Note, that the above matrix is antisymmetric, i.e. we have  $M^T = -M$  where  $^{*T}$  denotes the transpose of the matrix  $M$ .

If we now apply the concept of permutation test we notice that permuting observations and then computing the comparison matrix is quite the same as interchanging corresponding rows and columns in the above matrix Tab. 4.1. But interchanging columns does not affect the marginal sum vector at all and a permutation of the rows results in a corresponding permutation of the marginal sum vector.

Therefore, to derive the finite distribution of the statistic  $H^P$  we only have to apply one of the well-known algorithms, e.g. those for the distribution of the Wilcoxon-Mann-Whitney test, to the marginal sum vector, cf. STREITBERG and RÖHMEL (1986) for a review and NEUMANN (1988) for some ready to use computer programs.

But before we will give the solution for our example we have to make the above statements more precise thus proving that both statistics, the first,  $H^G$ , based only on the inter-sample comparisons the second,  $H^P$ , on both intra- and inter-sample comparisons are distributionally equivalent.

## 5. Distributional Equivalence

Let  $u^{ij}$  denote the matrix of the pairwise comparisons from the observations of sample  $i$  to those of sample  $j$ ,  $i, j = 1, 2$ , with respect to the comparison function (2.3). Note that by definition we have  $u^{ij} = -(u^{ji})^T$ . Then

$$(5.1) \quad u = \begin{pmatrix} u^{11} & u^{12} \\ u^{21} & u^{22} \end{pmatrix}$$

is a decomposition of the matrix of all intra- and inter-sample comparisons and the two statistics may be expressed in terms of matrix notation by

$$(5.2) \quad H^G = 1_{n_1}^T u^{12} 1_{n_2} = (1_{n_1}^T \ 0_{n_2}^T) u (0_{n_1}^T \ 1_{n_2}^T)^T$$

$$(5.3) \quad H^P = (1_{n_1}^T \ 0_{n_2}^T) u (1_{n_1}^T \ 1_{n_2}^T)^T$$

where  $1_n, 0_n$  denote the column vectors of length  $n$  with all components being equal to 1, 0, respectively. Evaluating (5.3) according to the decomposition (5.1) and taking into account the antisymmetry of  $u$ , especially  $u^{11} = -(u^{11})^T$ , we get

$$(5.4) \quad H^P = 1_{n_1}^T u^{11} 1_{n_1} + 1_{n_1}^T u^{12} 1_{n_2} = 0 + H^G$$

the general numerical equality of the two test statistics for a fixed combination of the samples. Next we have to show that the orbit of  $H^P$ , i.e. the set of all possible values under the group of permutations, is quite the same as that of  $H^G$ .

A permutation  $\pi \in \mathcal{S}_N$  may be represented by a matrix  $I_\pi$  with exactly one 1 in each row and each column and 0 else such that for the permutation  $\pi(x)$  of a vector  $x$  we have

$$(5.5) \quad \pi(x) = (x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_N})^T = I_\pi x$$

For the value of the statistic  $H^P$  under a permutation  $\pi$  of the vector  $x$  of the pooled  $N$  observations we thus have

$$(5.6) \quad H^P(\pi(x)) = (1_{n_1}^T \ 0_{n_2}^T) I_\pi u I_\pi^T 1_N \\ = 1_{\pi n_1}^T u_\bullet$$

where

$$1_{\pi n_1}^T = (1_{n_1}^T \ 0_{n_2}^T) I_\pi$$

is a vector of length  $N$  with exactly  $n_1$  entries 1 and 0 else and

$$u_\bullet = u I_\pi^T 1_N = u 1_N$$

is the vector of row sums. Thus under a permutation  $\pi$  the value of the statistic is given by summing up exactly  $n_1$  components of the vector of row sums. On the other hand, for the statistic based only on the inter-sample comparisons we have

$$(5.6) \quad H^G(\pi(x)) = (1_{n_1}^T \ 0_{n_2}^T) I_\pi u I_\pi^T (0_{n_1}^T \ 1_{n_2}^T)^T \\ = 1_{\pi n_1}^T u (1_N - 1_{\pi n_1})^T \\ = 1_{\pi n_1}^T u_\bullet - 1_{\pi n_1}^T u 1_{\pi n_1} \\ = 1_{\pi n_1}^T u_\bullet$$

which is exactly the same as for the other statistic. Here the last equality is a consequence of the antisymmetry of the matrix  $u$ . Hence for all permutations both statistics yield the same value and thus they are distributionally equivalent.

Since the statistic  $H^P$  is based on the permutations of a fixed vector, namely  $u_\bullet$ , the results from HAJEK (1969) are applicable. From this it follows that the variance under permutations is given by the quadratic form

$$(5.7) \quad S^2 = \frac{n_1 n_2}{(N-1) N} u_\bullet^T u_\bullet$$

where  $u_\bullet^T u_\bullet$  is just the sum of all squared row sums of the matrix of intra- and inter-sample comparisons. The root  $S$  may then be used for standardization in case of larger sample sizes.

It is an easy exercise to demonstrate the numerical equivalence of GEHAN'S formulas for the variance estimators and (5.7). But since this proof is somewhat lengthy it will be omitted here. Nevertheless, for the approximate versions of the respective test statistics the above formulation will be preferable because of its computational simplicity.

## 6. Completion of the Example

In our example from tab. 4.1 we have the row sum vector  $(-3, +2, -3, -10, -12, -7, +7, +2, -7, +7, +2, +10, +12)^T$ . Using the shift-algorithms, e.g. the program provided by NEUMANN (1988), or some other algorithm discussed by STREITBERG and RÖHMEL (1986) we get for the finite distribution of the statistic  $H^P$ .

Table 5. Finite distribution of the statistic  $H^P$  under permutations of the observations.

h	#	P (H = h)	Σ#	P (H ≤ h)	h	#	P (H = h)	Σ#	P (H ≤ h)
-42	1	.000583	1	.00058	0	19	.011072	877	.51107
-37	6	.003497	7	.00408	1	54	.031469	931	.54254
-33	6	.003497	13	.00758	2	43	.025058	974	.56760
-32	7	.004079	20	.01166	3	44	.025641	1018	.59324
-30	3	.001748	23	.01340	4	70	.040793	1088	.63403
-29	2	.001166	25	.01457	5	16	.009324	1104	.64336
-28	19	.011072	44	.02564	6	51	.029720	1155	.67308
-27	8	.004662	52	.03030	7	36	.020979	1191	.69406
-25	10	.005828	62	.03613	8	58	.033800	1249	.72786
-24	6	.003497	68	.03963	9	40	.023310	1289	.75117
-23	36	.020979	104	.06061	10	15	.008741	1304	.75991
-22	5	.002914	109	.06352	11	54	.031469	1358	.79138
-21	6	.003497	115	.06702	12	26	.015152	1384	.80653
-20	27	.015734	142	.08275	13	52	.030303	1436	.83683
-19	16	.009324	158	.09207	14	16	.009324	1452	.84615
-18	42	.024476	200	.11655	15	14	.008159	1466	.85431
-17	8	.004662	208	.12121	16	48	.027972	1514	.88228
-16	22	.012821	230	.13403	17	26	.009324	1530	.89161
-15	34	.019813	264	.15385	18	35	.020396	1565	.91200
-14	33	.019230	297	.17308	19	2	.001166	1567	.91317
-13	40	.023310	337	.19639	20	13	.007576	1580	.92075
-12	8	.004662	345	.20105	21	40	.023310	1620	.94406
-11	54	.031469	399	.23252	22	8	.004662	1628	.94872
-10	34	.019814	433	.25233	23	14	.008159	1642	.95688
-9	48	.027972	481	.28030	25	14	.008159	1656	.96503
-8	41	.023893	522	.30420	26	19	.011072	1675	.97611
-7	16	.009324	538	.31352	27	46	.003497	1681	.97960
-6	78	.045455	616	.35897	28	3	.001748	1684	.98135
-5	28	.016317	644	.37529	30	14	.008159	1698	.98951
-4	51	.029720	695	.40501	31	16	.003497	1704	.99301
-3	44	.025641	739	.43065	32	1	.000583	1705	.99359
-2	29	.016900	768	.44755	35	8	.004662	1713	.99825
-1	90	.052448	858	.50000	40	3	.001748	1716	1.00000

For the one-sided test decision we thus have a probability for a value less or equal than the one observed just by randomization alone of  $p_1 = .00758$ . For a two-sided test we have to add the corresponding tail probability at the opposite side of the distribution, resulting in a total probability of  $p_2 = .00758 + .00641 = .01399$ .

To demonstrate the computation of the variance necessary for the standardization of the statistic in case of large sample sizes we compute for our example

$$(6.1) \quad S^2 = \frac{6 \cdot 7}{12 \cdot 13} \quad (714) \\ = 192.23$$

from which we get as the statistic for an approximate test

$$(6.2) \quad T = \frac{H}{S} = -2.380$$

which if computed for really large samples, say  $n_1 \geq 20$ , has to be compared to the standard normal distribution. For our rather small example we would get as approximate tail probabilities  $p_{a,1} = .00866$  and  $p_{a,2} = .01732$ , respectively.

These values might give the impression that approximation behaves in a conservative way as compared to the exact randomization test. But this is not generally true since in other situations it also may be strongly anticonservative. In any case it is preferable to perform the exact test at least as far as it will be computationally feasible.

## References

- GEHAN, E. A. (1965a): A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203–223.
- GEHAN, E. A. (1965b): A generalized two-sample Wilcoxon test for doubly censored samples. *Biometrika* **52**, 650–653.
- HAJEK, J. (1969): *A Course in Nonparametric Statistics*. Holden-Day, San Francisco.
- HILGERS, R. (1988): Nonparametric Two-Sample Tests for General Clinical Data, to appear in *Biom. J.*
- NEUMANN, N. (1988): Some Procedures for Calculating the Distributions of Elementary Nonparametric Teststatistics. *Statist. Softw. Newsl.* **14** (in press).
- STREITBERG, B., RÖHMEL, J. (1986): Exact Distributions for Permutation and Rank Test: An Introduction to some Recently Published Algorithms. *Statist. Softw. Newsl.* **12**, 10–17.

Authors' address: Priv.-Doz. Dr. Dr. R. A. Hilgers, Dr. N. Neumann, Abteilung Medizinische Statistik der Georg-August-Universität Göttingen, Windausweg 2, D-3400 Göttingen

## Besondere Anlagentypen und ihre varianzanalytische Auswertung

Elvira Hofmann und Manfred Precht<sup>1</sup>

### Zusammenfassung

*Bei der Durchführung landwirtschaftlicher Feldversuche werden anstelle der klassischen Anlagen wie Blockanlage, Spaltanlage, Streifenanlage oder lateinisches Quadrat häufig andere Anlagentypen verwendet.*

*So ist es aus technischen Gründen oft erforderlich, einzelne Faktoren großflächig anzuwenden und beispielsweise ganze Streifen des Versuchsfelds auf die gleiche Art zu behandeln. Im Vordergrund dieser Arbeit stehen Varianten der Streifenanlage. Außerdem werden Kombinationen klassischer Versuchsanlagen, insbesondere Kombinationen mit lateinischen Quadraten, behandelt. Für die vorgestellten Anlagen werden die Varianzanalysetafeln ermittelt.*

### Summary

*In executing agricultural field experiments frequently there are used different types of designs instead of using one of the classical experimental designs like block-, split-plot-, split-block-plans or latin squares.*

*Because of technical reasons it is often required to apply some of the factors to big areas and for example to treat a whole stripe of the experimental field in the same way.*

*In this context variants of the split-block-design are in the foreground. Besides these also combinations with latin squares are presented.*

*In addition the tables for the analysis of variance for the designs just mentioned are developed.*

### 1. Einleitung

Ausgangspunkt der Untersuchung sind bekannte Anlagentypen, wie Block-, Spalt- oder Streifenanlagen. Im Gegensatz zur Blockanlage erlaubt die Spaltanlage die Anwendung eines Faktors auf größere Flächenstücke. Dies ist insbesondere dann von Vorteil, wenn die parzellenweise Anwendung einzelner Faktorstufen und Faktorstufenkombinationen auf technische Probleme stößt. So werden bei der Durchführung von landwirtschaftlichen Feldversuchen bestimmte Faktoren bevorzugt großflächig angewandt. Die Spaltanlage, Streifenanlage und Varianten davon erfüllen diese Anforderung und sind daher in der Praxis sehr beliebt.

---

<sup>1</sup> Die Autoren sind den Gutachtern für nützliche Hinweise sehr dankbar.

**2. Notation und Modelle**

Für die weiteren Betrachtungen wird die folgende Notation verwendet:

- $b$  Zahl der Blöcke
- $t_v$  Zahl der Stufen des  $v$ -ten Faktors
- $n$  Gesamtzahl der Beobachtungen

Grundlegend für die weiteren Ausführungen ist das allgemeine lineare Modell, das für einen zweifaktoriellen Versuch mit Wechselwirkung in seiner einfachsten Form wie folgt gegeben ist.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \varepsilon_{ijk} \tag{1}$$

Dabei ist  $\mu$  der allgemeine Mittelwert und  $\alpha_i (i = 1, \dots, b)$  der Blockeffekt.  $\beta_j (j = 1, \dots, t_1)$  und  $\gamma_k (k = 1, \dots, t_2)$  bezeichnen die Effekte der beiden Faktoren und  $(\beta\gamma)_{jk}$  deren Wechselwirkung.  $\varepsilon_{ijk}$  ist der  $(0, \sigma)$ -normalverteilte Zufallsfehler der Beobachtung  $y_{ijk}$ . Da jeder Randomisierungsschritt einen eigenen Fehlerterm beinhaltet, erhält man die Modelle für die nachfolgenden Anlagen, indem man, neben den zusätzlich auftretenden Effekten (z. B. Effekte der Zeile oder Spalte), die entsprechenden Fehlerterme in das Modell (1) aufnimmt. Mittelwerte werden wie üblich mit einem Querstrich gekennzeichnet, wobei ein Punkt als Index anzeigt, daß zur Mittelwertberechnung über diesen Index summiert wurde.

**3. Anlagentypen und Varianzanalysetafeln**

Die betrachteten Anlagen entstehen durch Modifizierung von Spalt- und Streifenanlagen sowie durch Kombinationen mit lateinischen Quadraten, wobei die Faktorstufen innerhalb eines Blockes mehrmals angewendet werden. Neben den nachfolgend vorgestellten Versuchsanlagen findet man andere Varianten bei (2) Federer, der sich mit zweifaktoriellen Spaltanlagen sowie Streifenanlagen beschäftigt, bei denen einer der Faktoren die Struktur eines lateinischen Quadrats besitzt. Beispiele zu vierfaktoriellen Spaltstreifenanlagen werden bei (4) Thöni behandelt. Weitere Beispiele für Kombinationen oder Überlagerungen klassischer Versuchstypen findet man auch in (3) und (6).

Einen für die Anwendung großflächiger Behandlungen nützlichen Anlagentyp stellt der folgende Versuchsplan dar. Die Anlage hat die Struktur der Streifenanlage, jedoch sind die Stufen der beiden Faktoren in jeweils  $m_v, v = 1, 2$  Streifen aufgespalten. Diese Streifen sind in jedem Block randomisiert angeordnet. Abbildung 3.1 zeigt ein entsprechendes Beispiel.

Für dieses Beispiel gilt:

$$b = 3, t_1 = 3, t_2 = 4, m_1 = 3, m_2 = 2, n = bt_1t_2m_1m_2$$

	3	1	2	1	3	2	4	4		1	3	2	3	2	4	1	4		2	4	3	2	1	4	3	1
C									C									C								
C									A									A								
A									A									C								
A									B									B								
B									C									A								
C									B									B								
B									C									B								
B									A									A								
A									B									C								

Abb. 3.1: Beispiel für Streifenvariante I

Betrachtet man den  $\nu$ -ten Faktor für sich alleine so zeigt sich, daß er in  $b$  randomisierten Blöcken auftritt, wobei jede Stufe  $m_\nu$ -mal pro Block beobachtet wird.

Tabelle 3.1 zeigt die Aufspaltung der Zahl der Freiheitsgrade bei der Varianzanalyse bezüglich des  $\nu$ -ten Faktors. Damit ergibt sich für die gesamte Anlage die in Tabelle 3.2 dargestellte Aufspaltung der Freiheitsgrade, wobei  $y_{ijklm}$  die Beobachtung darstellt mit  $i = 1, \dots, b, j = 1, \dots, t_1, k = 1, \dots, t_2, l = 1, \dots, m_1, m = 1, \dots, m_2$ .

Tab. 3.1: Freiheitsgrade bezüglich des  $\nu$ -ten Faktors

Variationsursache	Freiheitsgrade
Blöcke	$b-1$
innerhalb der Blöcke	
Faktor $\nu$	$t_\nu-1$
zwischen Beobachtungen	$(m_\nu-1)t_\nu b$
in den Stufen	} $(m_\nu-1)t_\nu b + (b-1)(t_\nu-1)$
Rest	
Gesamt	$bm_\nu t_\nu - 1$

Tab. 3.2: Freiheitsgrade bei der Varianzanalyse für Streifenvariante I

Variationsursache	Freiheitsgrade
Blöcke	$b-1$
1. Faktor	$t_1-1$
Rest a	$(b-1)(t_1-1) + t_1 b(m_1-1)$
2. Faktor	$t_2-1$
Rest b	$(b-1)(t_2-1) + t_2 b(m_2-1)$
Wechselwirkung	$(t_1-1)(t_2-1)$
Rest c	$(m_1 t_1 - 1)(m_2 t_2 - 1)b - (t_1 - 1)(t_2 - 1)$
Gesamt	$bm_1 m_2 t_1 t_2 - 1$

Tab. 3.3: SQ-Werte bei der Varianzanalyse für Streifenvariante I

Variationsursache	SQ-Werte	testen gegen
Blöcke	$t_1 t_2 m_1 m_2 \sum (\bar{y}_{i\dots} - \bar{y}_{\dots})^2$	
1. Faktor	$bt_2 m_1 m_2 \sum (\bar{y}_{\cdot j\dots} - \bar{y}_{\dots})^2$	Rest a
Rest a	$t_2 m_2 \sum (\bar{y}_{ij\cdot\cdot} - \bar{y}_{i\dots} - \bar{y}_{\cdot j\dots} + \bar{y}_{\dots})^2$	
2. Faktor	$bt_1 m_1 m_2 \sum (\bar{y}_{\cdot\cdot k\dots} - \bar{y}_{\dots})^2$	Rest b
Rest b	$t_1 m_1 \sum (\bar{y}_{i\cdot k\cdot m} - \bar{y}_{i\dots} - \bar{y}_{\cdot\cdot k\dots} + \bar{y}_{\dots})^2$	
Wechselwirkung	$bm_1 m_2 \sum (\bar{y}_{\cdot jk\dots} - \bar{y}_{\cdot j\dots} - \bar{y}_{\cdot\cdot k\dots} + \bar{y}_{\dots})^2$	Rest c
Rest c	$\sum (y_{ijklm} - \bar{y}_{ij\cdot\cdot} - \bar{y}_{i\cdot k\cdot m} + \bar{y}_{i\dots} - \bar{y}_{\cdot jk\dots} + \bar{y}_{\cdot j\dots} + \bar{y}_{\cdot\cdot k\dots} - \bar{y}_{\dots})^2$	
Gesamt	$\sum (y_{ijklm} - \bar{y}_{\dots})^2$	

Eine andere Variante einer Streifenanlage ergibt sich, wenn man ausgehend von Streifenvariante I bei beiden Faktoren die Streifen so ordnet, daß je  $t_\nu$  zu verschiedenen Stufen gehörige Streifen nebeneinander liegen. Seien  $p_l$  und  $p'_l$  ( $l = 1, \dots, m_1$ ) Permutationen von  $\{1, \dots, t_1\}$ , sowie  $q_\alpha$  und  $q'_\alpha$  ( $\alpha = 1, \dots, m_2$ ) Permutationen von  $\{1, \dots, t_2\}$ . Damit ergibt sich die in Abbildung 3.2 gezeigte Versuchsanlage.

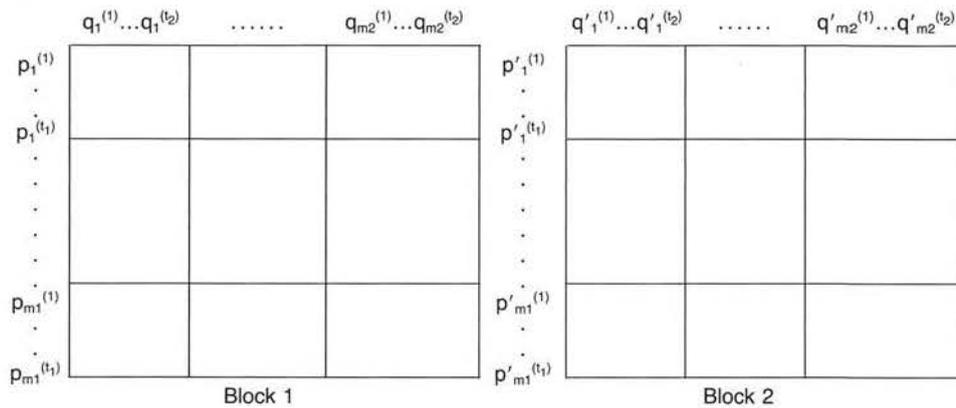


Abb. 3.2: Streifenvariante II

Innerhalb eines Blockes tritt jeder Faktor für sich allein betrachtet also genau in  $m_\nu$  vollständig randomisierten Unterblöcken auf. Damit ergibt sich für den  $\nu$ -ten Faktor die in Tabelle 3.4 dargestellte Aufspaltung der Zahl der Freiheitsgrade bei der Varianzanalyse. Analyse des Faktors  $\nu$  in insgesamt  $bm_\nu$  Unterblöcken:

Tab. 3.4: Freiheitsgrade für Faktor  $\nu$  bei Streifenvariante II

Variationsursache	Freiheitsgrade
Blöcke zwischen Unterblöcken in Blöcken	$b-1$
Faktor $\nu$	$(m_\nu-1)b$ } $bm_\nu-1$
Rest	$t_\nu-1$ $(t_\nu-1)(bm_\nu-1)$
Gesamt	$bm_\nu t_\nu-1$

Zusammenfassend erhält man die in Tabelle 3.5 angegebene Aufspaltung der Freiheitsgrade.

Tab. 3.5: Freiheitsgrade bei Streifenvariante II

Variationsursache	Freiheitsgrade
Blöcke in Blöcken zwischen Unterblöcken bezüglich Faktor 1	$b-1$
Faktor 1	$(m_1-1)b$
Rest a	$t_1-1$
in Blöcken zwischen Unterblöcken bezüglich Faktor 2	$(t_1-1)(bm_1-1)$
Faktor 2	$(m_2-1)b$
Rest b	$t_2-1$
Wechselwirkung	$(t_2-1)(bm_2-1)$
Rest c	$(t_1-1)(t_2-1)$
	$b(m_1 t_1 - 1)(m_2 t_2 - 1) - (t_1 - 1)(t_2 - 1)$
Gesamt	$bm_1 m_2 t_1 t_2 - 1$

Für  $m_\nu > 1$ ,  $\nu = 1, 2$  nimmt  $b$  in der Praxis oft den Wert 1 an. Ist  $m_1 = m_2 = 1$ , so ergibt sich die gewöhnliche Streifenanlage.

Beim Vergleich mit Streifenvariante I zeigt sich, daß sich die beiden Anlagen bezüglich der Freiheitsgrade von Rest c nicht unterscheiden, d. h., für den Test der Interaktion oder Wechselwirkung sind beide Anlagen gleichwertig. Für den Test der beiden Faktoren stehen jedoch bei Variante I mehr Nennerfreiheitsgrade zur Verfügung. Der höhere Grad an Systematik bei Variante II geht also zu Lasten der Freiheitsgrade für die Tests der beiden Faktoren.

Mit der eben aufgestellten Tafel der Varianzanalyse läßt sich auch der in Abbildung 3.3 gezeigte Versuchsplan auswerten.

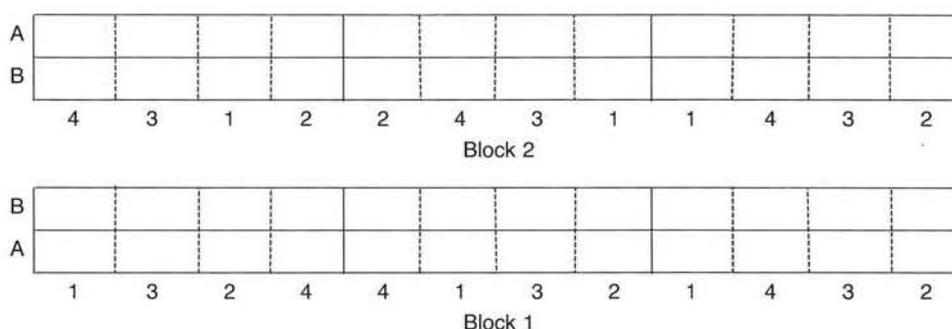


Abb. 3.3: Sonderfall der Streifenvariante II

Wie aus der Abbildung ersichtlich, ist dieser Anlagentyp nur ein Spezialfall von Streifenvariante II. Innerhalb eines Blockes gilt hier:  $m_1 = 1$ . Somit kann keine Variation in den Blöcken zwischen den Unterblöcken bezüglich Faktor 1 auftreten. Dies stimmt auch mit der

Tab. 3.6: SQ-Werte bei Streifenvariante II

Variationsursache	SQ-Werte	testen gegen
Blöcke	$t_1 t_2 m_1 m_2 \sum (\bar{y}_{i\dots} - \bar{y}_{i\dots\dots})^2$	
in Blöcken zwischen Unterblöcken		
bezüglich Faktor 1	$t_1 t_2 m_2 \sum (\bar{y}_{i\dots j\dots} - \bar{y}_{i\dots\dots})^2$	Rest a
Faktor 1	$b t_2 m_1 m_2 \sum (\bar{y}_{\cdot j\dots} - \bar{y}_{\cdot\dots\dots})^2$	Rest a
Rest a	$t_2 m_2 \sum (\bar{y}_{ij\dots} - \bar{y}_{i\dots\dots} - \bar{y}_{\cdot j\dots} + \bar{y}_{\cdot\dots\dots})^2$	
in Blöcken zwischen Unterblöcken		
bezüglich Faktor 2	$t_1 t_2 m_1 \sum (\bar{y}_{i\dots k\dots} - \bar{y}_{i\dots\dots})^2$	Rest b
Faktor 2	$b t_1 m_1 m_2 \sum (\bar{y}_{\cdot\dots k\dots} - \bar{y}_{\cdot\dots\dots})^2$	Rest b
Rest b	$t_1 m_1 \sum (\bar{y}_{i\dots k\dots m} - \bar{y}_{i\dots\dots m} - \bar{y}_{\cdot\dots k\dots} + \bar{y}_{\cdot\dots\dots})^2$	
Wechselwirkung	$b m_1 m_2 \sum (\bar{y}_{\cdot j k\dots} - \bar{y}_{\cdot j\dots} - \bar{y}_{\cdot\dots k\dots} + \bar{y}_{\cdot\dots\dots})^2$	Rest c
Rest c	$\sum (y_{ijklm} - \bar{y}_{ij\dots} - \bar{y}_{i\dots k\dots m} - \bar{y}_{\cdot j k\dots} + \bar{y}_{\cdot j\dots} + \bar{y}_{\cdot\dots k\dots} + \bar{y}_{i\dots\dots} - \bar{y}_{\cdot\dots\dots})^2$	
Gesamt	$\sum (y_{ijklm} - \bar{y}_{\cdot\dots\dots})^2$	

berechneten Zahl der Freiheitsgrade überein, die in diesem Fall Null ergibt. Gilt außerdem  $b = 1$ , so kann der erste Faktor nicht mit dem F-Test getestet werden, da in diesem Fall keine Nennerfreiheitsgrade zur Verfügung stehen.

Gelegentlich werden auch Anlagen verwendet, bei denen einer der Faktoren die Struktur eines lateinischen Quadrates aufweist. Eine Anlage dieser Art ist in Abbildung 3.4 dargestellt. Bei dieser Anlage handelt es sich um eine Variante der Spaltanlage. Die Hauptbehandlung liegt, wie bei der klassischen Spaltanlage in vollständig randomisierten Blöcken

vor, während jedoch die Unterbehandlungen innerhalb eines jeden Großteilstücks als vollständiges lateinisches Quadrat angeordnet sind.

2	1	3	3	1	2	2	3	1	3	2	1	Block 2
1	3	2	1	2	3	3	1	2	1	3	2	
3	2	1	2	3	1	1	2	3	2	1	3	
A			B			D			C			
3	1	2	2	1	3	1	2	3	3	2	1	Block 1
2	3	1	1	3	2	2	3	1	2	1	3	
1	2	3	3	2	1	3	1	2	1	3	2	
A			C			B			D			

Abb. 3.4: Spaltanlage, Unterbehandlungen in jedem Großteilstück als lateinisches Quadrat

Eine Anlage mit einer solchen Struktur findet man beispielsweise im Versuchsfeldführer 1982 des Staatsguts Roggenstein (Versuch 43). In diesem Versuch sollten unter anderem Ertrag und Qualität von Sommerweizen bei Herbst- und Frühljahrsaussaat untersucht werden. Die Hauptbehandlung Saatzeit lag also in zwei Stufen vor, die Unterbehandlungen waren fünf verschiedene Weizensorten, die innerhalb jeder Saatzeit als lateinisches Quadrat angeordnet waren.

Die Analyse der Großteilstücke liefert die folgenden Freiheitsgrade:

Tab. 3.7: Freiheitsgrade bei der Analyse der Großteilstücke

Variationsursache	Freiheitsgrade
Blöcke	$b-1$
Hauptbehandlungen	$t_1-1$
Fehler	$(b-1)(t_1-1)$
Gesamt	$bt_1-1$

Zur Analyse der Unterbehandlungen betrachtet man zunächst nur eine Stufe der Hauptbehandlung. Die Unterbehandlungen liegen somit in  $b$  lateinischen Quadraten vor. Die so erhaltenen Freiheitsgrade sind in der linken Spalte von Tabelle 3.8 aufgeführt. Die rechte Spalte zeigt die Ausweitung auf alle Stufen des ersten Faktors.

Tab. 3.8: Freiheitsgrade bei der Analyse der Kleinteilstücke

Variationsursache	Freiheitsgrade	
	für eine Stufe des 1. Faktors	für alle Stufen des 1. Faktors
Quadrate	$b-1$	$(b-1)t_1$
Zeilen in Quadraten	$(t_2-1)b$	$(t_2-1)bt_1$
Spalten in Quadraten	$(t_2-1)b$	$(t_2-1)bt_1$
Behandlungen	$t_2-1$	$(t_2-1)t_1$
Rest	$(t_2-1)(bt_2-b-1)$	$(t_2-1)(bt_2-b-1)t_1$

Damit erhält man bei der Varianzanalyse die in Tabelle 3.9 dargestellte Aufspaltung der Freiheitsgrade. In die Varianzanalysetafel geht die Beobachtung  $y_{ijkl(m)}$  ein, wobei die Indizes wie folgt zu verstehen sind: Block, Faktor 1, Zeile, Spalte und Faktor 2. Da die Beobachtung durch die ersten vier Indizes festgelegt ist, wird der 5. Index nur bei den SQ-Werten für die Unterbehandlung (Faktor 2) und die Wechselwirkung verwendet. Er ersetzt dann den Spaltenindex.

Tab. 3.9: Freiheitsgrade für Anlage aus Abbildung 3.4

Variationsursache	Freiheitsgrade
Blöcke	$b-1$
Hauptbehandlungen	$t_1-1$
Fehler a	$(t_1-1)(b-1)$
Zeilen in Quadraten	$(t_2-1)bt_1$
Spalten in Quadraten	$(t_2-1)bt_1$
Unterbehandlungen	$t_2-1$
Wechselwirkung	$(t_1-1)(t_2-1)$
Fehler b	$(t_2-1)(bt_2-b-1)t_1$
Gesamt	$bt_1t_2^2-1$

Tab. 3.10: SQ-Werte für die Anlage aus Abbildung 3.4

Variationsursache	SQ-Werte	testen gegen
Blöcke	$t_1t_2^2 \sum (\bar{y}_{i...} - \bar{y}....)^2$	Fehler a
Hauptbehandlungen	$bt_2^2 \sum (\bar{y}_{.j..} - \bar{y}....)^2$	
Fehler a	$t_2^2 \sum (\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}....)^2$	Fehler b
Zeilen in Quadraten	$t_2 \sum (\bar{y}_{ijk.} - \bar{y}_{ij..})^2$	
Spalten in Quadraten	$t_2 \sum (\bar{y}_{ij.l} - \bar{y}_{ij..})^2$	Fehler b
Unterbehandlungen	$bt_1t_2 \sum (\bar{y}_{...^{(m)}} - \bar{y}....)^2$	Fehler b
Wechselwirkung	$bt_2 \sum (\bar{y}_{.j^{(m)}} - \bar{y}_{.j..} - \bar{y}_{...^{(m)}} + \bar{y}....)^2$	Fehler b
Fehler b	$\sum (y_{ijkl} - \bar{y}_{ijk.} - \bar{y}_{ij.l} - \bar{y}_{.j^{(m)}} + \bar{y}_{ij..} + \bar{y}_{.j..})^2$	
Gesamt	$\sum (y_{ijkl} - \bar{y}....)^2$	

Verwendet man nur einen Block, so hat Fehler a null Freiheitsgrade und die Hauptbehandlungen können nicht getestet werden.

Auf ähnliche Weise läßt sich auch ein Versuch auswerten, bei dem sowohl die Hauptbehandlungen als auch die Unterbehandlungen innerhalb eines jeden Großteilstücks ein lateinisches Quadrat bilden, d. h. es liegt eine Schachtelung von lateinischen Quadraten vor. Dieser Anlagentyp eignet sich besonders für Versuche, bei denen Heterogenität in zwei Richtungen vermutet wird und außerdem ein Faktor nur auf größere Flächen anwendbar ist. Abbildung 3.5 zeigt eine solche geschachtelte Anlage.

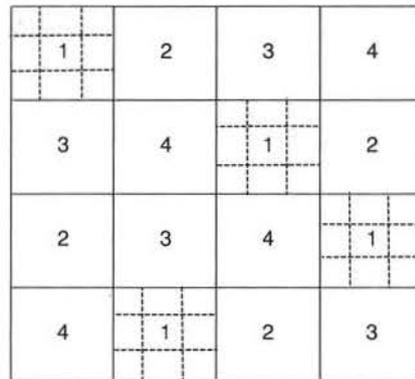


Abb. 3.5: Geschachtelte Spaltanlage, Haupt- sowie Unterbehandlungen in jedem Großteilstück als lateinisches Quadrat

Ähnlich wie bei der eben behandelten Anlage betrachtet man auch hier zunächst jeden Faktor für sich. Für die Hauptbehandlungen ergibt sich somit dieselbe Aufspaltung der Freiheitsgrade wie beim lateinischen Quadrat. Für die Unterbehandlungen betrachtet man zunächst nur eine Stufe der Hauptbehandlung. Der zweite Faktor kommt hierbei in  $t_1$  lateinischen Quadraten vor. Dies läßt sich genau wie bei der in Tabelle 3.8 dargestellten Tafel auf alle Stufen des ersten Faktors ausweiten. Nimmt man zusätzlich an, daß die Hauptbehandlung in  $b$  lateinischen Quadraten vorliegt, so erhält man die in Tabelle 3.11 gezeigte Aufspaltung der Zahl der Freiheitsgrade, wobei  $y_{ijk(l)mn(o)}$  die Beobachtung darstellt. Dabei bezeichnet der Index  $i$  das Quadrat,  $j$  und  $k$  die Zeile bzw. Spalte des Quadrats und  $m, n$  die Zeile bzw. Spalte im Großteilstück. Die Indizes  $(l)$  und  $(o)$  kennzeichnen die Stufen der beiden Faktoren. Die Mittelwerte sind analog zu Tabelle 3.9 zu verstehen.

Ein Beispiel für die Überlagerung eines lateinischen Quadrats mit einem streifenförmig angeordneten zweiten Faktor ist in Abbildung 3.6 dargestellt.

Tab. 3.11: Freiheitsgrade für Spaltanlage aus Abbildung 3.5

Variationsursache	Freiheitsgrade
Quadrate	$b-1$
Zeilen in Quadraten	$(t_1-1)b$
Spalten in Quadraten	$(t_1-1)b$
Hauptbehandlungen	$t_1-1$
Fehler a	$(t_1-1)(bt_1-b-1)$
Zeilen in Großteilstücken	$(t_2-1)bt_1^2$
Spalten in Großteilstücken	$(t_2-1)bt_1^2$
Unterbehandlungen	$t_2-1$
Wechselwirkung	$(t_1-1)(t_2-1)$
Fehler b	$(t_2-1)(bt_1t_2-bt_1-1)t_1$
Gesamt	$bt_1^2t_2-1$

Tab. 3.12: SQ-Werte für Spaltanlage aus Abbildung 3.5

Variationsursache	SQ-Werte	testen gegen
Quadrate	$t_1^2 t_2^2 \sum (\bar{y}_{i\dots\dots} - \bar{y}_{\dots\dots})^2$	Fehler a
Zeilen in Quadraten	$t_1 t_2^2 \sum (\bar{y}_{ij\dots\dots} - \bar{y}_{i\dots\dots})^2$	Fehler a
Spalten in Quadraten	$t_1 t_2^2 \sum (\bar{y}_{i\cdot k\dots\dots} - \bar{y}_{i\dots\dots})^2$	Fehler a
Hauptbehandlungen	$b t_1 t_2^2 \sum (\bar{y}_{\cdot\cdot(l)\dots\dots} - \bar{y}_{\dots\dots})^2$	Fehler a
Fehler a	$t_2^2 \sum (\bar{y}_{ijk\dots\dots} - \bar{y}_{ij\dots\dots} - \bar{y}_{i\cdot k\dots\dots} + \bar{y}_{i\dots\dots} - \bar{y}_{\cdot\cdot(l)\dots\dots} + \bar{y}_{\dots\dots})^2$	
Zeilen in Großteilstücken	$t_2 \sum (\bar{y}_{ijkm\dots\dots} - \bar{y}_{ijk\dots\dots})^2$	Fehler b
Spalten in Großteilstücken	$t_2 \sum (\bar{y}_{ijkn\dots\dots} - \bar{y}_{ijk\dots\dots})^2$	Fehler b
Unterbehandlungen	$b t_1 t_2 \sum (\bar{y}_{\dots\dots(o)} - \bar{y}_{\dots\dots})^2$	Fehler b
Wechselwirkung	$b t_1 t_2 \sum (\bar{y}_{\cdot\cdot(l)\cdot(o)} - \bar{y}_{\cdot\cdot(l)\dots\dots} - \bar{y}_{\dots\dots(o)} + \bar{y}_{\dots\dots})^2$	Fehler b
Fehler b	$\sum (y_{ijkmn} - \bar{y}_{ijkm\dots\dots} - \bar{y}_{ijkn\dots\dots} + \bar{y}_{ijk\dots\dots} - \bar{y}_{\cdot\cdot(l)\cdot(o)} + \bar{y}_{\cdot\cdot(l)\dots\dots})^2$	
Gesamt	$\sum (y_{ijkmn} - \bar{y}_{\dots\dots})^2$	

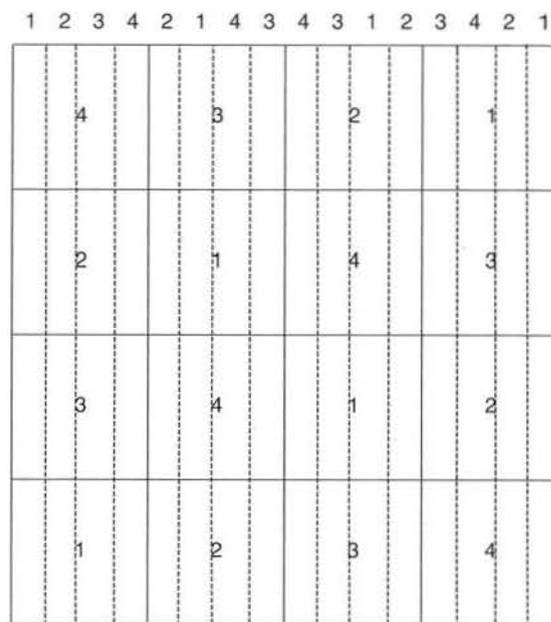


Abb. 3.6: Lateinisches Quadrat, streifenförmig überlagert

Die abgebildete Anlage wurde beispielsweise verwendet, um die Auswirkung von Fungizideinsatz und N-Düngung auf den Ertrag von Winterroggen zu untersuchen (Versuch 49, 1984 Roggenstein). Die Besonderheit dieser Anlage besteht darin, daß beide Faktoren wie lateinische Quadrate strukturiert sind, wobei beim zweiten Faktor die Zeilen nebeneinander liegen. Für die Auswertung betrachtet man zunächst beide Faktoren getrennt und faßt dann wieder zusammen. Mit  $t := t_1 = t_2$  ergibt sich die in Tabelle 3.13 gezeigte Aufspaltung der Freiheitsgrade (linke Spalte). Wären jedoch die Stufen der beiden Faktoren nicht nach der Struktur je eines lateinischen Quadrats angeordnet, sondern in den Spalten des ersten Quadrates und in den Zeilen des zweiten Quadrates beliebig randomisiert, könnte man den Plan wie eine Streifenanlage auswerten. Die entsprechenden Freiheitsgrade sind in der Tabelle 3.13 (rechte Spalte) zu finden.

Tab. 3.13: Vergleich der Freiheitsgrade für die Auswertung mit bzw. ohne Berücksichtigung der Quadratstruktur der Faktoren

Überlagerung zweier lateinischer Quadrate		Streifenanlage	
Variationsursache	FG	Variationsursache	FG
Zeilen	$t-1$	Blöcke	$t-1$
Spalten	$t-1$	1. Faktor	$t-1$
1. Faktor	$t-1$	Rest a	$(t-1)^2$
Rest a	$(t-1)(t-2)$	2. Faktor	$t-1$
2. Faktor	$t-1$	Rest b	$(t-1)^2$
Positionen (Spalten des 2. Faktors)	$t-1$	Interaktion	$(t-1)^2$
Rest b	$(t-1)(t-2)$	Rest c	$(t-1)^3$
Interaktion	$(t-1)^2$	Gesamt	$t^3-1$
Rest c	$(t-1)^3$		
Gesamt	$t^3-1$		

Ein Beispiel für die Kombination von Spaltanlage, Streifenanlage und lateinischem Quadrat ist in Abbildung 3.7 dargestellt. Der gezeigte Versuchsplan wurde in den Jahren 1984 und 1985 für Düngungsversuche mit Winterweizen verwendet (Versuchsfeldführer Roggenstein, 1984 und 1985, Versuche 45–48).

1. Faktor

	1	2	3	4	3	1	4	2	2	4	1	3	4	3	2	1
3																
2	1	2	1	2	2	2	1	1	2	2	1	1	1	2	1	2
1																
3																
2	2	1	2	1	1	1	2	2	1	1	2	2	2	1	2	1
1																

2. Faktor

Abb. 3.7: Kombination Spaltanlage – Streifenanlage – Lat. Quadrat

Der erste Faktor stellt ein lateinisches Quadrat dar, bei dem die Zeilen nebeneinander angeordnet sind. Der zweite Faktor ist mit soviel Wiederholungen (Blöcke) pro Grundplan angelegt, wie der dritte Faktor Stufen hat. Der zweite Faktor bildet mit Faktor 1 eine Streifenanlage, während Faktor 3 bezüglich Faktor 1 eine Unterbehandlung im Sinne der gewöhnlichen Spaltanlage darstellt. Auch hier werden zunächst nur Paare von Faktoren analysiert. Für die Faktoren 1 und 3 erhält man den Sonderfall einer Spaltanlage, bei der die Hauptbehandlung ein lateinisches Quadrat ist. Dagegen bilden die Faktoren 1 und 2 eine Streifenanlage-Variante, bei der der Faktor 2 in  $t_3 = 2$  Blöcken pro Grundplan auftritt. Somit laufen die Parzellen des ersten Faktors ganz durch die beiden Wiederholungen des zweiten Faktors und werden nicht getrennt für jede Wiederholung randomisiert. Daher werden die beiden übrigbleibenden Fehlerterme gepoolt und die restlichen Wechselwirkungstafeln gegen diesen gepoolten Wert geprüft. Eine Zusammenfassung dieser Varianzanalyse- tafeln ergibt die in Tabelle 3.14 gezeigte Aufspaltung der Freiheitsgrade, wobei  $g$  angibt, wie oft der Grundplan wiederholt wird (in Abb. 3.7 gilt  $g = 1$ ).

In Tabelle 3.14 ist die Variation zwischen den Blöcken in den Quadraten explizit angegeben. Dies ist bei der gleichmäßigen Verteilung der Stufen des dritten Faktors über die beiden Blöcke, wie sie in Abb. 3.7 gegeben ist, gerechtfertigt. Für den allgemeinen Fall, bei dem die Faktorstufen ungleichmäßig über diese Blöcke verteilt sind, gehen in die Unterschiede zwischen den Blöcken auch Unterschiede zwischen den Faktorstufen ein. Die Variation zwischen den Blöcken in den Quadraten ist also nicht frei von Unterschieden zwischen den Stufen des dritten Faktors angebar. Daher empfiehlt es sich, in dieser Situation die Variation zwischen den Blöcken in den Quadraten mit dem Fehler b zu poolen.

Eine Anlage wie in Abbildung 3.7 ist besonders sinnvoll, wenn man ein gleichbleibendes Gefälle von links nach rechts bzw. umgekehrt voraussetzt. Nach (1) Cochran & Cox, S. 118, kann man die Effekte des Gefälles am besten ausgleichen, wenn man ein lateinisches Quadrat verwendet, bei dem die Zeilen nebeneinander liegen, und dabei auch die Positionen innerhalb der jeweiligen Zeile berücksichtigt. Übertragen auf die in Abbildung 3.7 gezeigte Anlage bedeutet dies, daß man für die Tafel 3.14 voraussetzt, daß der Versuchsansteller die Art der Anlage bewußt gewählt hat, um ein Gefälle von links nach rechts bzw. umgekehrt auszugleichen. In diesem Fall kann man den Effekt der Position auch als differentiellen Effekt bezüglich der Zeile ansehen. Liegt jedoch kein derartiges Gefälle vor, so tritt im linearen Modell kein Positionseffekt auf. Die Variation zwischen den Spalten in den Quadraten muß dann mit Fehler a gepoolt werden.

Tab. 3.14: Tafel der Varianzanalyse für die Versuchsanlage aus Abbildung 3.7

Variationsursache	Freiheitsgrade
Quadrate (Grundpläne)	$g-1$
Zeilen in Quadraten	$(t_1-1)g$
Spalten in Quadraten	$(t_1-1)g$
Faktor 1	$t_1-1$
Fehler a	$(t_1-1)(t_1g-g-1)$
Blöcke in Quadraten	$(t_3-1)g$
Faktor 2	$(t_2-1)$
Fehler b	$(t_2-1)(t_2g-1)$
Faktor 3	$(t_3-1)$
Wechselwirkung $1 \times 3$	$(t_1-1)(t_3-1)$
Fehler c	$(t_3-1)(t_1g-1)t_1$
Wechselwirkung $1 \times 2$	$(t_1-1)(t_2-1)$
Wechselwirkung $2 \times 3$	$(t_2-1)(t_3-1)$
Wechselwirkung $1 \times 2 \times 3$	$(t_1-1)(t_2-1)(t_3-1)$
Fehler d (gepoolt)	$t_1t_3(t_1g-1)(t_2-1) + g-1-t_2(t_2g-1)$
Gesamt	$g^2t_1t_2t_3-1$

#### 4. Abschließende Betrachtungen

Allgemein lassen sich mehrfaktorielle Anlagen, die durch Überlagerung entstanden sind, dadurch auswerten, daß man zunächst die Tafeln der Varianzanalyse für die sich überlagernden Anlagen aufstellt und diese anschließend zusammenfaßt.

Vernachlässigt man die besondere Struktur einzelner Faktoren und wertet so aus, als ob eine der klassischen Anlagen vorläge, so hat dies oft zur Folge, daß z. B. Behandlungen getestet werden, für die genau genommen keine Nennerfreiheitsgrade für den F-Test zur Verfügung stehen.

Dennoch ist es manchmal erforderlich, statt der exakten Auswertung, d. h. Auswertung unter Berücksichtigung aller auftretenden Strukturen, eine Näherungslösung zu verwenden, die die Struktur der Anlage nur zum Teil berücksichtigt. Dies ist beispielsweise dann der Fall, wenn es einerseits die Bodengegebenheiten erfordern, Faktorstreifen wie Blöcke zu behandeln, und andererseits technische Gründe die streifenförmige Anordnung eines weiteren Faktors unumgänglich machen. So kann es aufgrund der Bodenbeschaffenheit notwendig sein, eine Anlage mit der Struktur der Streifenvariante II und  $m_1 = 1$ , sowie  $b = 1$  wie eine gewöhnliche Streifenanlage mit  $m_2$  Blöcken auszuwerten. Die Blockbildung fällt bei der Auswertung stärker ins Gewicht als die Systematik des 1. Faktors, der meist aus rein technischen Gründen streifenförmig auf alle Blöcke angewandt wird. Mit der Auswertung als klassische Streifenanlage ist nun ein F-Test für den ersten Faktor möglich, jedoch wird die Struktur der Anlage zum Teil vernachlässigt! Zu beachten ist außerdem, daß das allgemeine lineare Modell beim Übergang zur Näherungslösung verändert wird. In Fällen, in denen sowohl eine exakte als auch eine genäherte Auswertung vertretbar ist, ist das Versuchsziel mitentscheidend für die Wahl des Auswertungsschemas.

## 5. Literatur

1. COCHRAN, W. G., COX, G. M. (1968): Experimental design. Wiley.
2. FEDERER, W. T. (1977): Sampling, blocking and model considerations for split plot and split block designs, *Biom. J.* 19, S. 181–200.
3. HOFMANN, E. (1988): Zur Problematik der statistischen Auswertung landwirtschaftlicher Feldversuche. Dissertation TU München.
4. THÖNI, H. (1987): Mehrfaktorielle Versuche in Spaltstreifenanlagen, Vortragsmanuskript 33. Biometrisches Kolloquium, Trier.
5. Versuchsfeldführer: Gut Roggenstein, 1982–1985.
6. YATES, F. (1970): Experimental design, selected papers. Griffin, London.

## BUCHBESPRECHUNGEN/BOOK REVIEWS

### Programmieren in GAUSS

Eine Einführung in das Programmieren statistischer und numerischer Algorithmen

Von Dr. Ulrich KÜSTERS und Prof. Dr. Gerhard ARMINGER,

Gustav Fischer Verlag, Stuttgart · New York

1989. VIII, 315 S., 25 Abb., kt. DM 68,-. ISBN 3-437-40206-4

Das Buch ist didaktisch vorbildlich aufgebaut und führt anhand einfacher und komplexer Beispiele in die matrizenorientierte Programmiersprache GAUSS<sup>™</sup> ein. Mit dieser Sprache, die gleichzeitig eine sehr umfangreiche Bibliothek statistischer und mathematischer Routinen zur Lösung komplexer multivariater Fragestellungen enthält – z. B. zur Berechnung binärer Probit-Modelle –, steht dem Biometriker ein sehr effektives Hilfsmittel zur Verfügung, um in wenigen Befehlszeilen seine Probleme zu formulieren und auf IBM-kompatiblen PCs unter DOS zu lösen. Bisher stand dem GAUSS-Programmierer neben dem mit dem Programm mitgelieferten Handbuch keine systematische Einführung in die auf den ersten Blick komplex erscheinende Syntax zur Verfügung. Die beiden Autoren, die nicht nur über eigene große Erfahrung in der Programmierung mit GAUSS verfügen, sondern auch ständig mit konkreten multivariaten Anwendungsproblemen der Statistik konfrontiert sind, führen den an dieser Programmiersprache Interessierten hervorragend in die Syntax ein. Darüber hinaus geben sie wertvolle Hinweise auf mathematische Lösungswege und demonstrieren, wie einfach sich diese Gleichungssysteme mit GAUSS formulieren lassen. Gleichzeitig enthält das Buch wertvolle Hinweise, wie Programme übersichtlich zu strukturieren und zu erstellen sind.

Neben einer kurzen Einleitung gliedert sich das Buch in die Kapitel: Beispiele, Operatoren und Funktionen, Ein- und Ausgabe, Kontrollstrukturen und Unterprogramme, lineare Algebra mit GAUSS, Integration und nichtlineare Optimierung, Ergänzungen. In einem Anhang wird auf die mit dem RELEASE 2.0 zu erwartenden Neuerungen hingewiesen. Ein umfangreiches Literaturverzeichnis, auch zu mathematisch-statistischen Verfahren, rundet den guten Gesamteindruck ab.

Jedem etwas mathematisch vorgebildeten Anwender, der über die in den gängigen Statistikpaketen vorhandenen Routinen hinaus eigene Problemlösungen sucht und diese mit einem minimalen Programmieraufwand erstellen will, kann dieses Buch uneingeschränkt empfohlen werden.

Die Autoren sind zu ermutigen, eine ebenso klar geschriebene und gegliederte Anleitung für die Nutzung der in diesem System enthaltenen komplexen Graphik zu verfassen. (Berger, Hamburg)

M. S. KRAMER

### Clinical Epidemiology and Biostatistics

A Primer for Clinical Investigators and Decision-Makers

Springer, Berlin/New York, 1988

ISBN 3-540-18874-6. VII + 286 S. Gebunden. DM 76,00

Inhalt: Das Buch gibt eine Einführung in die Epidemiologie sowie in die Biostatistik, die sich für Medizin-Studenten und Kliniker eignet. Dabei ist Klinische Epidemiologie nicht als Einschränkung gegenüber der Allgemeinen Epidemiologie verstanden, sondern soll nur unterstreichen, daß sich das Buch besonders an Kliniker wendet. Im Teil über Biostatistik wird eine allgemeine Einführung gegeben und nicht etwa eine in spezielle statistische Methoden in der Epidemiologie.

Teil I Epidemiologische Forschungsansätze

1: Einführung

2: Messen

3: Raten

4: Epidemiologische Forschungsansätze: Ein Überblick

5: Verzerrungen

6: Kohortenstudien

7: Klinische Studien

8: Fall-Kontroll-Studien

9: Querschnittsstudien

**Teil II Biostatistik**

- 10: Einführung in die Statistik
- 11: Deskriptive Statistik
- 12: Hypothesentests und P-Werte
- 13: Inferenz-Statistik für stetige Variablen
- 14: Inferenz-Statistik für kategoriablen Variablen
- 15: Lineare Korrelation und Regression

**Teil III Spezielle Themen**

- 16: Diagnostische Tests
- 17: Entscheidungstheorie
- 18: Sterbekurvenanalyse
- 19: Kausalität

**Tafeln****Index**

Leserkreis: Kliniker, Medizin-Studenten, Statistiker (die sich mit Epidemiologie vertraut machen wollen).

Art der Darstellung: Elementare Einführung, wobei im Teil I Epidemiologie der Teil II Biostatistik noch nicht vorausgesetzt wird.

Bemerkungen: Das Buch gibt eine didaktisch gut aufbereitete Einführung in beide Themenkreise, wobei die Art der Darstellung die Denkprinzipien der Epidemiologie und Biostatistik betont und die mathematisch-statistischen Formalismen nur soweit wie unbedingt nötig heranzieht.

W. Lehmacher, Neuherberg

## PERSÖNLICHES



### Zum 70. Geburtstag von Professor F. X. Wohlzogen

Anlässlich des 70. Geburtstages von Professor FRANZ XAVER WOHLZOGEN und des 20jährigen Bestehens des Instituts für Medizinische Statistik der Universität Wien fand am 3. 4. 1989 ein Biometrisches Kolloquium der Wiener Sektion der Internationalen Biometrischen Gesellschaft, Region Österreich-Schweiz statt. Frau Professor L. HAVELEC würdigte dabei den Beitrag von Professor WOHLZOGEN für die Entwicklung der Biometrie in Österreich. Als junger Physiologe und British

Council Stipendiat wurde er erstmals 1952 in London mit statistischen Fragen konfrontiert, und zwar im Zusammenhang mit Problemen von Alles-oder-Nichts-Reaktionen. Mit dem ihm eigenen Blick für das Wesentliche erkannte er unmittelbar die große Bedeutung der Biostatistik für die medizinische Forschung und eignete sich in der Folge umfangreiche Kenntnisse des gesamten Fachgebietes an. Seine außerordentlichen intellektuellen Fähigkeiten setzten ihn bald in die Lage, nach neuen Wegen biometrischer Methoden zu suchen. Hier waren vor allem seine Arbeiten über sequentielle statistische Verfahren pionierhaft für die Biometrie. Das große Interesse für Entwicklungen auf dem Gebiet sequentieller Methoden besteht bis heute, und er hat es auch an die meisten seiner langjährigen Mitarbeiter weitergegeben. Darüber hinaus gilt sein Interesse vor allem der Planung von Experimenten, wobei er für diese Tätigkeit durch die profunden statistischen Kenntnisse und sein breites

medizinisches Wissen prädestiniert ist. Auch in dieser Hinsicht hat er versucht, seinen Mitarbeitern die Fähigkeiten zu vermitteln, den erforderlichen Kompromiß zwischen Theorie und Praxis zu suchen und die Verantwortung für angewandte Aufgaben der Biometrie zu übernehmen.

Professor WOHLZOGEN war wesentlich an der Gründungsversammlung der Region Österreich–Schweiz im Jahre 1961 beteiligt und prägte später auch als Präsident und Tagungsorganisator die positive Entwicklung der ROeS. Ende 1968 wurde das Wiener Institut für Medizinische Statistik und Dokumentation gegründet, zu dessen Leitung er berufen wurde. Heute ist dieses Wiener Institut eine Einheit mit 6 Mitarbeitern, das die vielfältigsten Aufgaben der Forschung und Lehre wahrnimmt. Zu den großen Projekten, an denen F. X. WOHLZOGEN zu seiner aktiven Zeit als Institutsleiter wesentlich beteiligt war, zählen die Organisation und Auswertung von Ringversuchen und das fast 10 Jahre dauernde interdisziplinäre Forschungsvorhaben zur Prüfung der Wirksamkeit von Fluoriden zur Kariesprophylaxe.

Frau Professor HAVELEC betonte auch das außergewöhnliche Verhältnis, das F. X. WOHLZOGEN zu seinen Mitarbeitern hatte. Der Rezensent kann dies aus eigener Erfahrung bestätigen; ungebrochen ist die Wertschätzung der Tätigkeit am Wiener Institut, die unter der Priorität kreativen Wirkens stand und formalen, bürokratischen Aspekten nur geringe Bedeutung einräumte. Professor WOHLZOGEN ist auch heute noch seinen ehemaligen Mitarbeitern eng verbunden, nachdem er 1986 in den Ruhestand getreten ist.

Neben Frau Professor HAVELEC trugen im Rahmen dieses Biometrischen Seminars noch drei weitere am Wiener Institut habilitierte Biometriker, V. SCHEIBER („Zwischenauswertungen: Gruppensequentielle statistische Verfahren“), P. BAUER („Sequentielle Meta-Analyse mit Lernen aus der Erfahrung“) und M. SCHEMPER („Inferenz und verallgemeinerte Kaplan-Meier Schätzer für standardisierte Prozesse“), vor, um quasi die wissenschaftliche Nachkommenschaft Professor WOHLZOGENS zu dokumentieren.

Die Zeitschrift gratuliert ihrem ehemaligen Mitarbeiter, der Rezensent gratuliert seinem ehemaligen Chef, dem er sich freundschaftlich verbunden fühlt.

P. BAUER, Universität zu Köln