

Ein einfaches wissensbasiertes System zur Auswahl statistischer Tests

Manhart, Klaus

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Manhart, K. (1989). Ein einfaches wissensbasiertes System zur Auswahl statistischer Tests. *Österreichische Zeitschrift für Statistik und Informatik (ZSI)*, 19(3), 289-300. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-52624>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Klaus Manhart

Ein einfaches wissensbasiertes System zur Auswahl statistischer Tests

Die vorliegende Arbeit behandelt das Problem der computergestützten Auswahl statistischer Verfahren am Beispiel statistischer Tests. Solche Systeme sind nützlich, da die in der empirischen Forschung verwendeten Statistik-Softwarepakete nicht vor der falschen Verwendung statistischer Standardverfahren schützen. Es wird ein einfaches wissensbasiertes System vorgestellt, das zugeschnitten ist auf Sozialwissenschaftler und in Interaktion mit dem Benutzer die Auswahl eines statistischen Tests vornimmt. Dem System unterliegen einfache Produktionsregeln, die aus einem Entscheidungsbaum konstruiert wurden. Hinweise für den Aufbau komplexerer statistischer Strategien werden am Schluß gegeben¹.

1. Einleitung

Die heutige Statistik ist gekennzeichnet durch eine Vielzahl teilweise sehr komplexer statistischer Verfahren, die ohne den Einsatz moderner Rechenanlagen kaum mehr durchführbar wären. Entsprechende Statistik-Software ist gegenwärtig von Unterprogramm-Bibliotheken bis hin zu integrierten Paketen in vielen Formen verfügbar.

Integrierte statistische Softwarepakete wie SPSS, SAS, BMDP oder GLIM sind insbesondere in den empirischen Wissenschaften weit verbreitete Analyseinstrumente: sie haben exzellente numerische Fähigkeiten, sind relativ leicht handhabbar und decken die in der empirischen Forschung anfallenden statistischen Probleme weitgehend ab.

Die leichte Bedienbarkeit dieser Pakete unterstützt allerdings auch deren unreflektierten Gebrauch. So haben Statistiker immer wieder Substanzwissenschaftler in ihrer unkritischen – oft schlicht falschen – Verwendung statistischer Standardverfahren kritisiert (vgl. z.B. das Zitat von Chambers in HAHN 1985, Seite 1). Speziell naive Benutzer, also Anwender mit nur geringen Kenntnissen in Statistik, sind anfällig für die falsche Verwendung statistischer Verfahren, da die gängigen Softwarepakete blind alle Daten in syntaktisch richtiger Form akzeptieren, ohne semantische Fehler erkennen zu können: "The statistical procedures in SPSS have little ability to distinguish between proper and improper applications of the statistical techniques. They are basically blind computational algorithms that apply their formulas to whatever data the user enters" (NIE et al. 1975, Seite 3).

So kommt es in der Praxis immer wieder vor, daß Verfahren angewendet werden, deren Annahmen auf die Daten gar nicht zutreffen: "One can score a nominal variable using the

¹ Der Autor bedankt sich sehr herzlich bei dem anonymen Gutachter für wertvolle Literaturhinweise und Vorschläge zum Aufbau der Arbeit.
Der Beitrag war in etwas anderer Form Teil einer Jahresarbeit am Institut für Höhere Studien, Wien.

integers and then go ahead and calculate means, standard deviations, carry out t-tests and so on without so much as a murmur from the program ... In brief, simple programs for calculators and computers and complex packages for computers assume the user knows what he is doing" (HAND 1984, Seite 353).

Neben der fast unbeschränkten Möglichkeit, beliebige Verfahren auf beliebige Daten anzuwenden, bietet die heutige Statistik-Software darüber hinaus kaum Hilfe bei der Interpretation der gelieferten Ergebnisse. Derzeitigen Statistikpaketen fehlt somit eindeutig eine "intelligente" Komponente, die die Benutzer führt, sie vor Mißbrauch warnt und hilft, Ergebnisse zu interpretieren.

Die Forderung nach "intelligenter" Statistik-Software wurde mit der Ausbreitung des Artificial Intelligence Paradigmas immer vehementer erhoben (vgl. etwa HAHN 1985). Insbesondere wissensbasierte Systeme wurden als prädestiniert betrachtet, die oben beschriebenen Probleme zu bewältigen und den Benutzer vor dem Mißbrauch statistischer Verfahren zu warnen, ihn zu führen und bei der Ergebnisinterpretation zu helfen.

Diese wissensbasierten Systeme könnten Aktivitäten übernehmen, die spezifisch für einen Statistiker sind: das Wissen eines Statistikers wird in einem Computerprogramm repräsentiert und kann auf konkrete Fälle angewendet werden. Bei der Kommunikation mit dem Benutzer fragt das System nach Daten (oder besser: "Wissen"), zieht aus diesem Wissen und dem implementierten Fachwissen Schlüsse und trifft schließlich begründbare Entscheidungen. Wissensbasierte Systeme stellen somit Expertenwissen für einen größeren Benutzerkreis zur Verfügung und sind zeitlich und räumlich jederzeit verfügbar. Da die Domäne wissensbasierter Systeme heutzutage eingeschränkt ist auf Expertenwissen, werden solche Systeme von den meisten Autoren synonym als Expertensysteme bezeichnet; wir schließen uns dieser Terminologie an. Für eine Einführung in die Expertensystemtechnologie vgl. z.B. HAYES-ROTH/WATERMAN/LENAT 1983.

In der Statistik können Expertensysteme eingesetzt werden (HAUX 1986b, HAHN 1985):

- in verschiedenen Phasen der statistischen Analyse (Benutzerführung zu Informationsquellen, statistische Probleme definieren und Modelle auswählen, angemessene Methoden vorschlagen, Analyse und Interpretation);
- in verschiedenen Bereichen statistischer Methoden (Kontingenztafelanalyse, Regressionsanalyse etc.);
- in verschiedenen Anwendungsgebieten (Biologie, Medizin, Psychologie, Soziologie etc.).

Darüber hinaus können diese auf verschiedene Benutzergruppen zugeschnitten sein:

- statistisch naive Benutzer (wie statistisch nicht ausgebildete Mediziner);
- statistisch ausgebildete Substanzwissenschaftler;
- Fachstatistiker.

Es gäbe somit eine breite Palette von Einsatzmöglichkeiten für statistische Expertensysteme; ein Überblick über das Gebiet Expertensysteme und Statistik sowie die Vorstellung verschiedener realisierter Systeme findet sich in GALE (1986), HAUX (1986a), DE ANTONI/LAURO (1986).

2. Problemstellung und Systemanforderung

HAND (1984) sieht die gegenwärtigen Einsatzmöglichkeiten statistischer Expertensysteme vor allem in zwei unterschiedlichen Phasen statistischer Analyse: in der Phase der Auswahl einer statistischen Technik und in der Phase der Anwendung dieser Technik. Ein Beispiel für ein System, das in der Phase der Anwendung eines statistischen Verfahrens eingesetzt werden könnte, ist etwa REX (PREGIBON/GALE 1984), das statistisch naive Benutzer bei der Durchführung einer Regressionsanalyse unterstützt. Der Vorschlag, daß die Durchführung einer Regressionsanalyse überhaupt ein geeignetes statistisches Verfahren ist, wäre die Aufgabe eines Expertensystems in der 1. Phase.

Die vorliegende Arbeit behandelt das Problem der Auswahl einer statistischen Technik am Beispiel statistischer Tests. Die automatische Auswahl statistischer Tests erscheint als ein ausgezeichnetes Anwendungs- und Einsatzgebiet für wissensbasierte Systeme:

- einen für eine Fragestellung geeigneten Test zu finden ist ein alltägliches Problem in der Forschungspraxis;
- es existiert eine Vielzahl von Tests für die unterschiedlichsten Probleme unter unterschiedlichen Anwendungsvoraussetzungen;
- das Auffinden eines den Daten angemessenen statistischen Tests ist nicht trivial;
- es gibt (formalisierbares) Expertenwissen, das zum Teil explizit in Lehrbüchern abgelegt ist, zum Teil aber auch intuitive Erfahrungskomponenten berücksichtigt.

Eine computergestützte Hilfe bei der Testauswahl hätte insbesondere bei statistisch naiven Benutzern ein Ansteigen der Qualität statistischer Analysen zur Folge. Da statistisches Testen ein umfangreiches und komplexes Gebiet ist und jeder Benutzerkreis spezifische Anforderungen stellt, wurde dieser auf eine klar abgegrenzte Gruppe eingengt. Es sollte ein Prototyp erstellt werden für Psychologen und Sozialwissenschaftler, die mit geringen Statistikkenntnissen einen Test durchführen wollen und das Programmpaket SPSS benutzen. Die Benutzung von SPSS ist nicht entscheidend, das System orientiert sich aber an den hier zur Verfügung stehenden Tests.

Das Programm sollte die üblichen parametrischen und parameterfreien Tests kennen, welche in den einschlägigen Lehrbüchern und in knapper Form im SPSS-Handbuch beschrieben sind. Als Referenzliteratur wurden insbesondere die beiden Bücher von RÜGER (1985) und SIEGEL (1976) verwendet.

Als Ergebnis sollte das System – in Interaktion mit dem Benutzer – einen Test vorschlagen und eine Rechtfertigung geben, warum dieser bestimmte Test abgeleitet wurde. Das Wissen um die Struktur der Daten soll nicht von den Daten selbst abgeleitet, sondern vom Benutzer

erfragt werden. Es sollten einfache, für den Benutzer verständliche Fragen gestellt werden und die Antworten sollten hinreichend sein, einen Test vorschlagen zu können. Die Durchführung des Tests selbst ist nicht mehr Aufgabe des Systems, sondern kann mit einem beliebigen numerischen Paket wie SPSS erfolgen.

Wir nennen dieses System TESTSEL (TEST SElection). Unter Benutzung der oben angegebenen "Nomenklatur" soll TESTSEL also verwendet werden:

- in der Phase der "Selektion einer statistischen Technik";
- im Bereich statistische Methoden: "Testen statistischer Hypothesen";
- im Anwendungsgebiet: "Psychologie und Sozialwissenschaften".

TESTSEL ist ferner zugeschnitten auf statistisch naive Benutzer mit Basiskenntnissen in Statistik.

3. Grundlegende Ideen

Die grundlegende Annahme, die hinter TESTSEL liegt, ist darin begründet, daß die Wahl eines statistischen Tests als ein regelgeleiteter Vorgang betrachtet wird, der mit Hilfe eines (großen) Entscheidungsbaumes repräsentiert werden kann. Solche Entscheidungsbäume sind als Anleitung für Benutzer auch in der Literatur weit verbreitet (vgl. etwa für parameterfreie Tests SIEGEL 1976).

Im vorliegenden Fall bildet die Wurzel des Entscheidungsbaums das Problem der Auswahl eines Tests, das in der nächsten Ebene in verschiedene Unterprobleme zerfällt, welche vom Benutzer zu bestimmen sind. Solche Unterprobleme sind etwa

- Tests für Lageparametervergleich im 1-Stichprobenfall
- Tests für Lageparametervergleich im 2-Stichprobenfall
- Tests auf Verteilung Güte der Anpassung etc.

Nachdem klar ist, um welchen Problemfall es sich handelt (dem Benutzer stehen umfangreiche Informationstexte zur Verfügung), verzweigt der Baum weiter in die "Wissensstrukturen" um die Daten, welche jeweils die Knoten des Baumes bilden. In Abhängigkeit vom Problemfall sind dies etwa die Knoten:

- Stichprobengröße(n)
- Skalenart(en)
- Kenntnis der Varianz(en) etc.

Die Blätter des Baumes bilden dann den mit dem gewählten Pfad determinierten Test.

Der Entscheidungsbaum bildet die Grundlage für die Ableitung der Regeln. Eine solche Regel hat die allgemeine Form

Wenn A und B und ..., dann T,

wobei A, B Elementaraussagen über die Daten (Datenbankitems) und T den gewählten Test symbolisieren. In der Terminologie des Entscheidungsbaumes formuliert, bildet die dann-

Komponente T das Blatt und die wenn-Komponenten die von der Wurzel zum Blatt führenden Knoten des Pfades.

Die folgende Regel (R) wird als Beispielregel im Fortlauf des Textes weiter verwendet; (R) dürfte selbsterklärend sein:

(R) WENN

- a) der Mittelwert einer Stichprobe mit einem theoretischen Wert verglichen werden soll

UND

- b) die Daten metrisch skaliert sind

UND

- c) die Stichprobengröße kleiner als 40 ist

UND

- d) bekannt ist, daß die Grundgesamtheit normalverteilt ist

UND

- e) die Varianz der Grundgesamtheit bekannt ist

DANN (Schlußfolgerung)

- f) Nimm den einfachen Gauss-Test.

Aus dem Entscheidungsbaum können dann alle einen Test determinierenden Regeln abgeleitet werden.

4. Wissensrepräsentation

Die Auffassung, wonach die Auswahl eines statistischen Tests ein regelgeleiteter Prozeß ist, deckt sich mit der heute üblichen Wissensrepräsentationsform in Expertensystemen. In den meisten Expertensystemen wird das Wissen in Form von sogenannten Produktionsregeln repräsentiert, die Teil eines Produktionssystems sind. HAND (1985) schlägt diese Repräsentationsart auch für die Auswahl statistischer Techniken vor.

Produktionssysteme bestehen aus einer Menge von Produktionsregeln, einer dynamischen Datenbank (Fakten), auf welche die Regeln angewendet werden und einem Inferenz- oder Ausführungsmechanismus, der die Regelanwendung steuert.

Produktionsregeln sind Regeln der oben angegebenen Art aus einer logischen Konjunktion von Datenbankitems und der Aktion (Konklusion) als Konsequenz. Die Datenbankitems (Fakten) sind entweder vom Benutzer eingegebene Antworten auf entsprechende Fragen oder sie werden vom System abgeleitet. Der Inferenzmechanismus "matcht" (vergleicht) das

Antecedens mit den Daten der dynamischen Datenbank. Wenn das Antecedens erfüllt ist, also alle Bedingungen der Regel (R) (a) – e)) in der Datenbank stehen, dann "feuert" die Regel und führt die Aktion f) aus, d.h. schlägt den Gauss-Test vor.

Der Inferenzmechanismus kann hierbei grundsätzlich vorwärts (Forward-Chaining) oder rückwärts (Backward-Chaining) arbeiten. Beim Forward-Chaining geht man von den Fakten aus, wendet die Regeln an, deren Vorbedingungen durch die Fakten erfüllt sind und fügt die Aktion zu den anderen Fakten hinzu. Dies wird wiederholt, bis keine Regel mehr anwendbar ist. Beim Backward-Chaining, das in TESTSEL realisiert ist, wird von der Aktion als Hypothese ausgegangen und es wird versucht, sämtliche Bedingungen der Regel nacheinander zu beweisen. In der Beispielregel (R) wird der einfache Gauss-Test als hypothetisch anzuwendender Test betrachtet und es wird geprüft, ob die Bedingungen erfüllt sind. In TESTSEL sind die Bedingungen so implementiert, daß sie im Fall der Regel (R) in folgender Reihenfolge geprüft werden:

- Testproblem 1-Stichprobenfall
- Intervallskala
- Stichprobengröße < 40
- Grundgesamtheit normalverteilt
- Varianz der Grundgesamtheit bekannt.

Ist eine Bedingung nicht erfüllt, werden die anderen nicht weiter überprüft und es wird der nächste hypothetische Test betrachtet.

5. Struktur des Systems

Das vorliegende System beinhaltet die folgenden Komponenten, die als Minimalforderungen an ein wissensbasiertes System gelten (vgl. HAYES-ROTH/WATERMAN LENAT 1983) und die das oben angeführte Auswahlverfahren durchführen:

- eine Wissensbasis, bestehend aus Produktionsregeln und dynamischer Datenbank
- ein Inferenzmechanismus, der die Abarbeitung der Regeln gewährleistet
- eine Dialogkomponente, die den Dialog mit dem Benutzer regelt
- eine Erklärungskomponente, die dem Benutzer zeigt, durch welche Fakten und Regeln eine Folgerung zustande kam.

TESTSEL wurde in Prolog implementiert. Prolog unterstützt direkt das Programmieren mit Regeln und Fakten in Backward-Chaining. Das klassische Prolog-Lehrbuch ist CLOCKSIN/MELLISH (1984).

Der statische Teil der Wissensbasis besteht aus einer Menge von ca. 40 Regeln. In Anbetracht dessen, daß das "berühmteste" Expertensystem MYCIN – ein System zur Unterstützung bei der Diagnose bakterieller Infektionen – ca. 450 Regeln enthält und größere Systeme mehrere 1.000, ist TESTSEL also wirklich klein. In der dynamischen Datenbank werden die Antworten des Benutzers auf entsprechende Fragen des Systems gespeichert; die Antworten werden nach jedem Durchlauf gelöscht.

Als Inferenzmechanismus wird ohne Modifikation die Strategie des Prolog-Beweismechanismus übernommen. Prolog kann als automatischer Theorembeweiser betrachtet werden, der aus Fakten und Regeln neue Fakten ("Theoreme") beweist; der Beweismechanismus beruht auf einer speziellen Methode, dem Resolutionsverfahren von Robinson. Prolog arbeitet die Regeln sequentiell von oben nach unten ab: es beginnt mit dem ersten hypothetischen Test und versucht von links nach rechts die Bedingungen zu beweisen. In der Regel (R) wäre dies – nachdem der Mittelwertvergleich für den 1-Stichprobenfall vom Benutzer festgelegt wurde – die Bedingung, daß die Daten metrisch skaliert sind. Da das System in der dynamischen Datenbank keine Antwort über Skalenarten vorfindet, stellt es an den Benutzer eine entsprechende Frage. Gibt der Benutzer ein, daß die Daten metrisch skaliert sind, wird dieser Eintrag gespeichert und es wird die nächste Bedingung überprüft. Angenommen aber, der Benutzer gibt Rangskalenniveau ein, so trägt das System als Skalenart Ordinalskala ein, geht zum nächsten hypothetischen Test über (der aktuelle kann ja nicht zutreffen wegen des geforderten Intervallskalenniveaus) und versucht, diesen herzuleiten. Angenommen, in diesem wird nur Rangskalenniveau gefordert, so findet das System diesen Eintrag bereits und braucht den Benutzer nicht noch einmal bezüglich der Skalenart zu konsultieren.

Der Dialog des Benutzers mit dem System erfolgt fast ausschließlich menügesteuert, da der Benutzer immer nur die Wahl zwischen einigen wenigen Alternativen hat.

Die Erklärungskomponente liefert eine Begründung für die Auswahl eines Tests. Die Erklärung erfolgt ganz einfach als wenn-dann-Regel nach Vorlage der obigen Regel (R), wobei die Bedingungen direkt aus der Wissensbasis übernommen werden.

Hinzugefügt wurde ferner eine Option, sich nach der Auswahl eines Tests Literatur angeben zu lassen, in welcher der Test detailliert beschrieben ist und die Option, sich den ausgewählten Test kurz am Bildschirm beschreiben zu lassen.

Aus Benutzersicht präsentiert sich TESTSEL mit einem Auswahlmenü, bei dem das spezielle Testproblem anzugeben ist. Hierzu steht ebenfalls eine Option zur Verfügung, mit der detailliertere Informationen zu den einzelnen Wahlmöglichkeiten abrufbar sind.

6. Probleme beim "Knowledge Engineering"

Beim Aufbau größerer wissensbasierter Systeme geht man in der Regel so vor, daß man neben der Konsultation von Fachliteratur einen Experten eines Fachgebietes (ev. auch mehrere) interviewt und versucht, die spezifischen Kenntnisse, die der Experte zur Lösung des Problems einsetzt, in Erfahrung zu bringen. Den Prozeß der Extrahierung des Wissens aus der Literatur und dem menschlichen Experten, der Umsetzung dieses Wissens in formale Fakten und Regeln und deren Implementierung bezeichnet man als Knowledge Engineering.

Im vorliegenden Fall wurde das Wissen ausschließlich Lehrbüchern entnommen. Die größte Schwierigkeit war dabei, daß nicht immer die Bedingungen angegeben waren, die einen Test determinieren. Die Kriterien für die Testauswahl waren nicht nur lückenhaft, sondern in manchen Fällen auch widersprüchlich.

All dies spricht nicht gegen den Einsatz von Expertensystemen, sondern dafür: in typischen Expertendomänen ist Lehrbuchwissen lückenhaft und das Wissen verschiedener Experten in manchen Fällen widerspruchsvoll.

Dennoch können Experten Probleme ihres Fachgebiets lösen, weil sie über einen reichen Erfahrungshintergrund verfügen und implizites Erfahrungswissen angesammelt haben, das nicht in den Lehrbüchern zu finden ist. Wenn auch das Vorgehen verschieden ist und das Wissen verschiedener Experten manchmal widerspruchsvoll, kommen verschiedene Experten meist zum gleichen Ergebnis aufgrund unterschiedlich gemachter Erfahrungen und Problemlösungsstrategien.

Ein menschlicher Experte hilft auf alle Fälle – wenn auch aus seiner Sicht – Lücken, Unschärfen und Widersprüche von Lehrbuchtexten zu beseitigen und dem System eine echte Erfahrungskomponente zu geben. Zweifellos ist der fehlende menschliche Experte ein Manko von TESTSEL.

7. Aufbau komplexerer statistischer Strategien

Das vorliegende Regelsystem in seiner schlichten Struktur kann als einfaches hypothetisch-deduktives Vorgehen betrachtet werden: es identifiziert Zielzustände (Hypothesen) und versucht, die Bedingungen, die zu diesen Zielzuständen führen, abzuleiten.

HAND (1984, 1985) tritt dafür ein, dieses hypothetisch-deduktive Vorgehen stärker auf die Mechanismen **menschlicher** Statistikexperten auszurichten. Der Ansatz, die kognitiven Prozesse menschlicher Experten nachzubilden, wird von vielen Autoren vertreten und hätte u.a. den Vorteil, daß die Akzeptanz solcher Systeme bei potentiellen Benutzern größer wäre.

HAND (1984) verifizierte seine Vorstellung, wonach statistisches Problemlösen ein (komplexer) hypothetisch-deduktiver Prozeß ist anhand von Tonbandprotokollen bei statistischen Konsultationen (eine Beschreibung verschiedener Sessions findet sich in HAND (1984, Seite 358 – 361). Er stellte hierbei starke Parallelen fest zwischen dem Prozeß, wie Statistiker eine Methode auswählen und dem Prozeß, wie Humanmediziner Diagnosen erstellen.

Statistisches Rasonieren läßt sich danach als hypothetisch-deduktiver Prozeß auffassen, bei dem eine anfängliche – auf wenig Information beruhende – "Diagnose" generiert wird und durch weitere Fragen und Tests bestimmte Hypothesen verstärkt und andere widerlegt werden.

Danach könnte ein Computerprogramm einige wahrscheinliche statistische Methoden bzw. in unserem Fall statistische Tests generieren und dann durch Backward-Chaining bestimmen, welcher adäquat ist.

Konkret liest Hand's Programm eine Kurzbeschreibung des Projektes bzw. Problems in Form eines kleinen Textes ein und versucht mit elementaren Methoden des "Natural Language Processing" (Mustererkennung) statistisch relevante Schlüsselwörter bzw. Phrasen zu entdecken. Diese Schlüsselwörter veranlassen das System, bestimmte potentielle Verfahren zu generieren, die anschließend tiefer erforscht werden. Enthält der Eingabetext also bei-

spielsweise das Wort "Gruppen", dann könnten die folgenden Ziele aktiviert werden, wobei die Zahlen in Klammern die Priorität andeuten sollen (HAND 1984, Seite 365):

- Varianzanalyse (1)
- t-Test (2)
- Multivariate Varianzanalyse (2)
- Kovarianzanalyse (2)
- Nicht-parametrischer Gruppenvergleich (2)
- χ^2 -Test (3)
- Log-lineares Modellieren (3).

Werden auf dieser Stufe unangemessene Techniken vorgeschlagen, ist dies weiter nicht nachteilig, da – mit gezielten Fragen an den Benutzer – diese mit der Zeit vom System verworfen werden und die Aufmerksamkeit auf die Alternativen gelenkt wird. Dies ist die Grundstruktur von Hands Programm; für nähere Details vgl. HAND (1984).

Bei einer Diskussion mit zwei Fachstatistikern wurde dem Autor ein dem Ansatz von Hand analoges Verfahren vorgeschlagen, nach dem das System nach einigen einleitenden Fragen eine Liste statistischer Tests vorschlägt und durch weitere gezielte Fragen bestimmte Tests ausschließt und die Evidenz für andere erhöht².

Dieses Vorgehen hat u.a. den Vorteil, daß der Benutzer nicht präzise wissen muß, wo genau das Problem liegt. In TESTSEL muß der Benutzer zumindest schon wissen welche Testproblematik vorliegt. In der Praxis ist es aber oft so, daß nur diffuse Vorstellungen vorliegen, die erst nach und nach konkreter werden. Auf diesen Sachverhalt verweist auch HAHN (1985).

Mit reinen Produktionssystemen sind die Grenzen dieses am menschlichen Vorgehen orientierten Ansatzes allerdings bald erreicht: "... we felt that a simple unordered set of antecedents did not provide sufficient structure to represent statistical methods ... This feeling is in tune with the suggestion that a cookbook approach to statistics is inadequate" (HAND 1985, Seite 10).

Die Erfahrungen mit reinen Produktionssystemen veranlaßten Hand zur Forderung nach Systemen mit mehr Struktur im Antecedens von Regeln, mehr Struktur in der Beziehung zwischen Regeln und einer Betonung auf breitere Mustererkennungsprozesse. HAND (1985) stellt ein hochstrukturiertes Expertensystem zur Auswahl statistischer Techniken vor, das diesen Kochbuchansatz vermeidet und als Resultat seiner Erfahrung mit einfachen Ansätzen betrachtet werden kann. Für ausführliche Details vgl. HAND (1985).

Ein anderer Aspekt zum Aufbau komplexerer statistischer Strategien betrifft die Klärung von Punkten (Begriffen), die statistisch naive Benutzer unter Umständen nicht verstehen. Die obige Regel (R) fragt z.B. den Benutzer, ob die Daten intervallskaliert vorliegen. Es kann sein, daß der Benutzer nicht weiß oder sich nicht sicher ist, was eine intervallskalierte Variable ist. Einfache Expertensysteme folgen der Konvention, nach der in solchen Fällen Erklärungen in Form mehr oder weniger langer Texte ausgegeben werden, beginnend mit

² Die Anregungen verdanke ich den Herren Alt und Polasek (Institut für Statistik und Informatik, Universität Wien bzw. Universität Basel).

„eine intervallskalierte Variable ist ...“. (In TESTSEL ist diese Möglichkeit zwar nicht realisiert, kann aber leicht hinzugefügt werden.) HAND (1985) kritisiert dies als Lehrbuchansatz, der für Expertensysteme wenig geeignet erscheint, und argumentiert dafür, daß das System selbst durch näheres Befragen des Benutzers die Skalenart erforschen sollte. Ist der Benutzer also unfähig, die Skalenart anzugeben, dann versucht das System selbst, es herauszufinden, indem es Regeln benutzt, für die die Aktion lautet, daß es sich um eine Intervallskala handelt.

Ein dritter Aspekt schließlich betrifft die Verknüpfung der symbolischen Expertensystem-Welt mit der traditionellen numerischen Software-Welt. Ein Nachteil von „Stand-alone“-Expertensystemen ist u.a., daß Benutzer Dateneigenschaften fehlerhaft wahrnehmen, was dann als Faktenwissen in die Wissensbasis eingeht. CARLSON/HEUCH (1986) treten dafür ein, den wissensbasierten Ansatz mit den traditionellen numerischen Softwarepaketen zu koppeln und so die Vorteile beider „Welten“ zu verbinden. Die Autoren stellen ein Expertensystem EXPRESS (EXPert system Relying on Existing Statistical Software) vor zur Testauswahl beim Testen auf gleiche Lageparameter im 2-Stichprobenproblem und verwenden als numerische Software BMDP.

EXPRESS beruht auf der Grundidee, daß statistisches Testen ein iterativer Prozeß ist, bei dem auf jeder Stufe statistischer Analyse Informationen aus den numerischen Daten kombiniert werden mit Kontextwissen, bevor zur nächsten Stufe weitergegangen wird.

Ein Statistiker mit minimalem Vorwissen beginnt z.B. beim Test auf gleiche Lageparameter im 2-Stichprobenproblem damit, daß er einfache statistische Kennwerte wie Schiefe oder Kurtosis betrachtet um zu bestimmen, ob Normalverteilung eine vernünftige Anpassung ist; alternativ kann er einige einfache Plots betrachten. Verwirft er die Normalverteilungsannahme, kann er etwa die Effekte verschiedener Transformationen untersuchen. Er kann dann einen t-Test anwenden auf das Original oder die transformierten Daten, mit oder ohne Annahme gleicher Varianzen. Die Entscheidung über die Varianzen kann auf einem F-Test basieren. Falls alle Versuche zur Stützung der Normalverteilungsannahme fehlschlagen, können die Daten einem nicht-parametrischen Verfahren unterzogen werden wie dem Mann-Whitney-Test.

EXPRESS simuliert dieses Vorgehen, wobei es auf Möglichkeiten achtet, die ein Standardpaket wie BMDP anbietet. Das System ist fähig, die Ausführung des Pakets zu initiieren, den Output zu interpretieren, die wesentlichen Ergebnisse festzuhalten und dann über den nächsten Schritt zu entscheiden. Die wichtigste Neuigkeit dieses Ansatzes liegt aber in dem wiederholten Gebrauch des Pakets, wobei Entscheidungen auf früheren Ergebnissen beruhen.

Allerdings dürfen die Schwierigkeiten bei der Koppelung der numerischen Statistik-Softwarewelt mit der symbolischen Expertensystemwelt nicht unterschätzt werden. Bei der Kombination sind grundsätzlich zwei Ansätze möglich:

- man schreibt ein Expertensystem in einer symbolisch orientierten Programmiersprache wie Lisp oder Prolog mit einem Interface zu unabhängigen statistischen Routinen;
- man baut das Expertensystem innerhalb der Umgebung des Statistik-Paketes, d.h. verwendet numerisch orientierte Sprachen.

EXPRESS ist in Fortran programmiert "because of external restrictions". Dies erleichtert die Verbindung zum numerischen BMDP, erschwert aber die Implementierung von Expertensystemtechniken. Die Autoren selbst meinen, daß viele praktische Probleme leichter gewesen wären, wenn sie eine symbolische Programmiersprache verwendet hätten – was allerdings wieder die Verbindung zur numerischen Welt erschwert hätte.

Dennoch ist unter Statistikern eher die Tendenz feststellbar, in der bekannten Umgebung zu arbeiten, als auf Werkzeuge aus der Artificial Intelligence umzusteigen. Ein extremes Beispiel liefert DARIUS (1986), der eine Expertensystem-Shell vollständig in SAS entwickelte.

8. Schlußbetrachtung

Es wurde ein einfacher Prototyp eines wissensbasierten Systems zur Auswahl statistischer Tests präsentiert, der als praktisches Werkzeug sicherlich nur mit erheblichen Verfeinerungen einsetzbar ist. Die Erfahrung bei der Entwicklung des Systems lehrt, daß auf die Konsultation eines Statistikexperten nicht verzichtet werden sollte. Zur Vermeidung eines Kochbuchansatzes und der Erhöhung der Akzeptanz sollten statistische Expertensysteme am menschlichen Vorgehen ausgerichtet sein und vor allem eine reichhaltigere Struktur aufweisen. Ein Vergleich mit der Literatur zeigt, daß solche Systeme – wenn auch mit hohem Aufwand – mit den gegenwärtigen Techniken realisierbar sind.

Für die nächste Zukunft sind entsprechend komplexe, aber bequeme Entwicklungswerkzeuge zu erwarten, die den Aufwand erträglicher gestalten könnten.

Literatur

- CARLSEN, F., HEUCH, I. (1986);
 "Express – An Expert System Utilizing Standard Statistical Packages", in: De Antoni Lauro (1986), 265 – 270.
- CLOCKSIN, W.F., MELLISH, C.S. (1984);
 Programming in Prolog, Berlin: Springer.
- DARIUS, P.L. (1986);
 "Building Expert Systems with the Help of Existing Software: An Example", in: De Antoni Lauro (1986), 271 – 282.
- DE ANTONI, F., LAURO, N. (ed.) (1986);
 Proceedings of COMPSTAT 1986, Wien: Physica.
- GALE, A. (ed.) (1986);
 Artificial Intelligence and Statistics, Reading, Mass.: Addison-Wesley.
- HAHN, G.J. (1985);
 "More Intelligent Statistical Software and Statistical Expert Systems: Future Directions", in: The American Statistician, 39, 1, 1 – 8.

- HAND, D.J. (1984);
"Statistical Expert Systems: Design", in : The Statistician, 33, 351 – 369.
- HAND, D.J. (1985);
"Choice of Statistical Technique", in: Proceedings of the 45th ISI-Session, Vol. LI, Book 3, Amsterdam.
- HAUX, R. (ed.) (1986a);
Expert Systems in Statistics. Selected Papers from a Workshop, Stuttgart: Fischer.
- HAUX, R. (1986b);
"Prologue", in: Haux (1986a), 1 – 6.
- HAYES-ROTH, F., WATERMAN, D.A., LENAT, D.B. (ed.) (1983);
Building Expert Systems, Reading, Mass.: Addison-Wesley.
- NIE, N.H. et al. (1975);
Statistical Packages for the Social Sciences, New York: MacGraw Hill.
- PREGIBON, D., GALE, W.A. (1984);
"REX: An Expert System for Regression Analysis", in Havranek. T. et al., Proceedings of COMP-STAT 1984, Wien: Physica, 242 – 248.
- SIEGEL, S. (1976);
Nicht-parametrische statistische Methoden, Frankfurt/Main: Fachbuchhandlung für Psychologie.
- RÜGER, B. (1985);
Induktive Statistik, Oldenbourg: München - Wien.