

ANWENDUNGSBEZOGENER KRITERIENKATALOG
ZUR BEURTEILUNG VON
STANDARDANWENDUNGSSOFTWARE FÜR DIE
AUSWERTUNG VON MARKTFORSCHUNGSDATEN

Norbert FRÖMMER *)
Karlheinz ZÜGNER

Forschungsbericht/
Research Memorandum No. 163

Mai 1981

*) Scholaren an der Abteilung Betriebswirtschafts-
lehre/OR, Institut für Höhere Studien, Wien

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich der Autoren und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.

Z u s a m m e n f a s s u n g

Die Bedeutung der Marktforschung nimmt in der heutigen informationsbetonten Zeit ständig zu. Die optimale Nutzung erhobener Daten ist nur nach sorgfältiger Auswertung möglich.

Der erarbeitete Beurteilungskatalog, der bei der Auswahl von Standardanwendungssoftware zur Auswertung von Marktforschungsstudien die Entscheidungsgrundlage verbessern soll, enthält zahlreiche Methoden, die bei realistischer Sample-Größe ohne Computerunterstützung nicht einsetzbar sind. Deshalb wird der Softwaremarkt kurz beleuchtet und die im ISIS Software Report 1/1981 zusammengestellte Software des Bereiches "Marktforschung und Statistik" gegenübergestellt.

Anhand von Beispielen wird deutlich gemacht, daß die Mehrzahl der bisher veröffentlichten Kriterienkataloge relativ global und daher wenig praktikabel bei der Problemlösung erscheinen. Der auf fünf Seiten zusammengefaßte Kriterienkatalog betont dagegen Auswertungsmethoden (Häufigkeitsauswertungen, Tests, Korrelationen, Regressionen, Varianz-, Faktoren-, Diskriminanz- und Clusteranalyse), das Daten- und Datei-Handling sowie die Gestaltung der Ausgabe. In Exkursen werden Itemüberprüfung und Zeitreihenanalyse behandelt. Nach einer Kurzcharakteristik der in den Katalog aufgenommenen Verfahren wird auf 16 Seiten der Detail-Kriterienkatalog präsentiert, der die problemorientierte Beurteilungsgrundlage im Rahmen einer Nutzwertanalyse bilden könnte.

A b s t r a c t s

An important feature of marketing research is the optimum evaluation of data. Sophisticated methods cannot be applied without the help of computers. Since IBM's decision to unbundle software and hardware the number of software packages has steadily increased. Therefore we give a short survey of the market of application software and analyse some packages which specialize in the evaluation of market research data.

Most of the criteria lists published to compare packaged software are very general. The catalogue presented in this paper concentrates on methods of data evaluation (frequencies, tests, correlations, regressions, variance analysis, factor analysis, discriminant analysis and cluster analysis) as well as on data handling and output production. We also mention time series analysis and item tests.

The tabular summary is followed by a short description of the methods and a 16-page catalogue of detailed criteria.

VORWORT

Die vorliegende Arbeit baut auf einer Seminararbeit auf, die im Anschluß an die Vorlesungsreihe "Aufbau betrieblicher Informationssysteme", die o.Univ.Prof.Dr. Hans R. Hansen, Wirtschaftsuniversität Wien, im Januar 1981 als Gastprofessor am Institut für Höhere Studien und Wissenschaftliche Forschung Wien hielt, verfaßt wurde.

Die beiden Verfasser - zu dieser Zeit Scholaren der Abteilung Betriebswirtschaftslehre und Operations Research am IHS - versuchten, einen Beurteilungskatalog für Standardanwendungssoftware im Bereich der Auswertung von Marktforschungsdaten zu erstellen.

Im Gegensatz zum überwiegenden Teil der bisher veröffentlichten Kriterienkataloge, welche durchwegs globale und somit wenig praktikable Beurteilungspunkte betonen, liegt das Schwergewicht der hier vorgestellten Zusammenstellung auf Methoden, die eine verbesserte Datenanalyse zulassen. Dennoch wird die Bedeutung der Dateneingabe und -manipulation sowie die Gestaltung der EDV-Ausgabe nicht vernachlässigt.

Für wertvolle Hinweise danken Herrn Professor Hansen die Verfasser

Mag. Norbert Frömmer
und
Mag. Karlheinz Zügner

INHALTLICHE GLIEDERUNG

	Seite
1. Zur Bedeutung der Marktforschung	2
2. Der Markt für Standardanwendungssoftware	4
2.1 Allgemeiner Überblick	4
2.2 Standardanwendungssoftware für die Marktforschung	8
3. Kriterien für die Auswahl von Standard- anwendungssoftware	13
4. Kriterienübersicht	18
4.1 Methoden	18
4.2 Daten- und Datei-Handling	21
4.3 Ausgabegestaltung	21
5. Erläuterung der Methoden	23
5.1 Häufigkeitsauswertungen	23
5.2 Tests	24
5.3 Korrelationen	25
5.4 Regression	27
5.5 Varianzanalyse	29
5.6 Faktorenanalyse	32
5.7 Diskriminanzanalyse	34
5.8 Clusteranalyse	36
Exkurs 1:Überprüfung von Fragebogenitems	38
6. Detail-Kriterienkatalog	39
6.1 Methoden	39
6.1.1 Häufigkeitsauswertungen	39
6.1.2 Tests	41
6.1.3 Korrelationen	42
6.1.4 Regression	44
6.1.5 Varianzanalyse	45
6.1.6 Faktorenanalyse	46
6.1.7 Diskriminanzanalyse	47
6.1.8 Clusteranalyse	48
6.2 Daten- und Datei-Handling	49
6.3 Ausgabegestaltung	51
Exkurs 1:Überprüfung von Fragebogenitems	53
Quellenverzeichnis	55

1. ZUR BEDEUTUNG DER MARKTFORSCHUNG

Marktforschung wird definiert als die "systematisch betriebene Erforschung der Märkte eines Unternehmens, insbesondere die Analyse der Fähigkeit dieser Märkte, Umsatz hervorzubringen" (Meffert, S.148).

Daraus resultiert die Aufgabe, Informationen über die Umwelt eines Unternehmens, das sind die jeweiligen Absatz- und Beschaffungsmärkte, bereitzustellen (siehe Abb.1).

Gewinnung externer Informationen	
Erforschung der Beschaffungsmärkte, z. B. hinsichtlich Lieferanten, Beschaffungswegen, Lieferfristen, Transportmitteln, Preisen und Qualitäten	Erforschung der Absatzmärkte, z. B. hinsichtlich Marktpotential, Marktvolumen und Marktanteil, Konkurrenz, Bedarfsstruktur, einzelner Elemente des Marketing-Mix (Produkt, Preis, Werbung, Distribution)
Marktforschung	

Abb.1: Gewinnung unternehmensexterner Informationen als Aufgabe der Marktforschung (Berekoven, S.20)

In modernen Unternehmenskonzepten bilden Informationen die Grundlage für die Steuerung eines Betriebes ("To manage a business well is to manage its future - and to manage its future is to manage information" (M.Harper, zitiert in Kotler, S.419). Die Güte jeder Entscheidung basiert unabhängig ihres Geltungsbereiches auf der Qualität und Quantität der Informationen, die ihr zugrundeliegen.

Hierin liegt die wachsende Bedeutung der Marktforschung im Bereich des Marketing ("Marketing information has become the critical input into effective marketing ..." / Kotler, S.443) und der gesamten Unternehmung begründet.

Die Auswertung von Marktforschungsdaten hat großteils einen so hohen Grad an Komplexität erreicht, daß eine Marktforschungsstudie ohne EDV-Unterstützung unmöglich erscheint. Daher erscheint die Erstellung eines Kriterienkataloges für Standardanwendungssoftware zur Auswertung von Marktforschungsdaten gerechtfertigt und geeignet, den Erfolg eines Unternehmens zu steigern, da bei der Investition in ein Standardanwendungssoftware-Paket potentielle Fehlentscheidungen vermieden werden können.

Wie im dritten Abschnitt noch genauer erläutert wird, konzentrieren wir uns in dem erarbeiteten Beurteilungskatalog in erster Linie auf Methoden, die für die Auswertung von Marktforschungsdaten herangezogen werden können bzw. sollen, wenn das Datenmaterial den diesbezüglichen Anforderungen entspricht. Der Einsatz derartiger Methoden innerhalb der Marktforschungsfunktion Auswertung ist bei realistischen Sample-Größen ohne Computerunterstützung praktisch nicht durchführbar. Bevor aber mit großem Programmieraufwand "das Rad ein zweites Mal" erfunden wird, sollte man sich einen Überblick über die für diese Funktion angebotene Standardanwendungssoftware verschaffen.

2. DER MARKT FÜR STANDARDANWENDUNGSSOFTWARE

2.1 ALLGEMEINER ÜBERBLICK

Eine konstante Verbesserung des Preis-/Leistungsverhältnisses bei EDV-Hardware, einhergehend mit einer stetigen Zunahme der Personalkosten im EDV-Bereich hat zu einer Umkehr des Verhältnisses der Kosten von Hardware zu Software geführt (Abb.2).

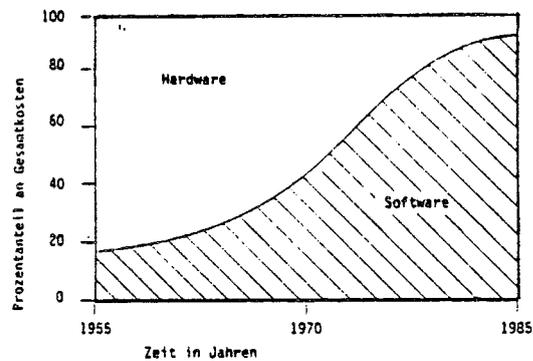


Abb.2: Die Entwicklung der Hardware- und Software-Kosten als Prozentanteil an den Gesamtkosten (Frank, S.19)

Die größten Rationalisierungsmöglichkeiten im EDV-Bereich ergeben sich folglich bei Software. Durch den Einsatz von mehrfach verwendbarer Standardanwendungssoftware, so werden fertige Programme definiert, die auf Allgemeingültigkeit und vielseitige Nutzung hin angelegt sind, kann

- Kostenersparnis (die "economies of scale" entstehen durch die Verteilung der Entwicklungskosten auf eine Vielzahl von

- Anwendern und die Kostenreduktion aufgrund der Erfahrungskurve),
- Zeitgewinn (eine unternehmensinterne Neuentwicklung dauert im Normalfall länger als die Implementation von geeigneter Standardanwendungssoftware),
 - die Kompensation vorhandener Personalengpässe und der Ausgleich fehlenden Know-Hows erreicht werden (vgl. Hansen 1981, S.3).

In der BR Deutschland wurden 1978 auf dem Anwendungssoftwaremarkt 1,2 Mrd DM umgesetzt, davon entfielen 1/3 auf Standardanwendungssoftware. In Österreich wird das Volumen des Standardanwendungssoftwaremarktes derzeit auf 500 Mio öS geschätzt. Wachstumschancen ergeben sich primär aus der Tatsache, daß bei EDV-Großanwendern noch immer 75% und bei Basisdatenverarbeitungsanwendern 45% der verwendeten Software eigenprogrammiert wurden (vgl. Infratest/MBP, S.23). Trotzdem wird die Entwicklung dieses Marktes unterschiedlich beurteilt, wobei sich euphorische Beurteilungen der Zukunft (siehe beispielsweise o.V., Computer Magazin 8/1980, S.37ff) und Voraussagen über krisenhafte Entwicklungen (siehe zB Kanngiesser, Online 12/1980, S.1006ff und Komarnicki, Online 1/1978, S.42ff) die Waage halten.

1980 wurden 2294 Standardsoftware-Programme von Software-Häusern (75% der Anbieter) und Hardware-Herstellern (25%) angeboten, von denen 662 neu auf dem Markt platziert wurden, während 309 aus dem Angebot ausschieden (entspricht insgesamt einer Steigerungsrate von 18,2% von Juli 1979 bis Juli

1980). Abbildung 3 zeigt die Verteilung des Gesamtangebotes mit Stand Juli 1980 (in Klammern die Zahlen des Vorjahres) auf der Basis des ISIS Software Reports, Ausgabe 2/1980.

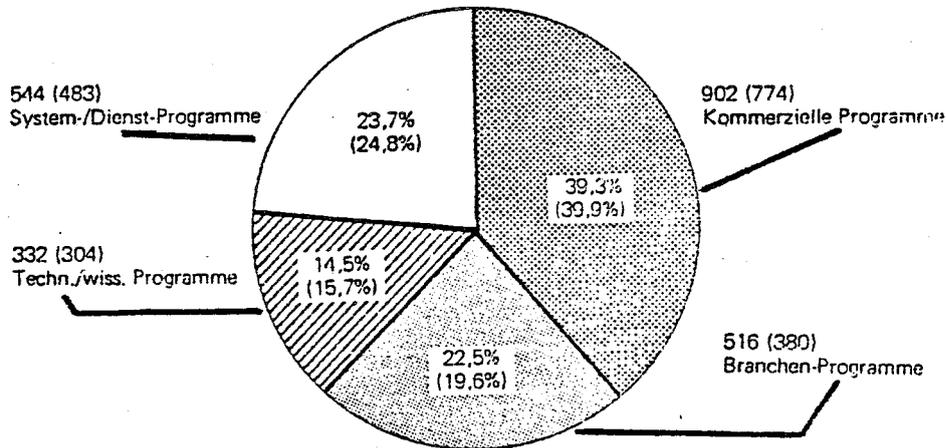


Abb.3: Struktur des Gesamtangebotes an Standardanwendungssoftware

Die hier relevanten kommerziellen Programmpakete für die diversen betrieblichen Funktionsbereiche erreichten einen Anteil von 40% des Gesamtangebotes und sind damit die wichtigste Sparte auf dem Standardanwendungssoftware-Markt, jedoch weist der Bereich der Branchenprogramme stärkere Wachstumszahlen auf.

Die Angebotsschwerpunkte auf dem Teilmarkt der "Kommerziellen Programme" lassen sich der folgenden Tabelle entnehmen:

822 Kommerzielle Programme			
Finanz-/Rechnungswesen		Fertigungsplanung	
Finanzbuchhaltung		Fertigungssteuerung	57
Rechnungswesen	125	Marketing/Vertrieb	
Kreditorenbuchhaltung	15	Vertriebsplanung	
Debitorenbuchhaltung	18	Vertriebssteuerung	
Anlagenbuchhaltung		Prognose-Systeme	
Anlagenrechnung	47	Zeitreihenanalyse	
Unternehmensplanung/ Finanzplanung/ Planungsorachen	33	Auftragsbearbeitung	
Informationssysteme		Fakturierung	
DBS Base Services	3	Abrechnungssysteme	
Kostenrechnung	89	Fakturierung	
Personalwesen		Adressverwaltung	
Lohn-/Gehaltsrechnung	104	Direktwerbung	
Massenkommunikation	29	Marktforschung/Statistik	
Lagerhaltung	14	Textverarbeitung	
Beschaffung/Einkauf	11	Projektplanung, Projektüberwachung	
		Netzplantechnik	
		34	

Tab.1: Verteilung der kommerziellen Programme auf funktionale Unternehmensbereiche (Hansen / Maier, S.61)

Die praktische Implementierung dieser Programmpakete in Unternehmen ist noch relativ gering. Nur etwa 1/4 der angebotenen Standardanwendungssoftware erreicht mehr als 20 Installationen und gar nur 10% wurden mehr als 50mal bei verschiedenen Unternehmen implementiert.

	Installationsgruppe*)						TS/o. Ang.
	0	1	2	3	4	5	
Kommerzielle Programme Basis: 850 Programme	7,7	27,9	19,5	13,6	14,5	10,1	6,7
Branchen-Programme Basis: 433 Programme	17,3	38,8	12,9	11,1	7,4	6,5	6
Techn./wiss. Programme Basis: 331 Programme	6,9	35,6	12,1	5,4	7,3	6,4	26,3
System-/Dienst-Programme Basis: 526 Programme	7,8	20,9	13,3	8,2	10,5	34	5,3

*) Es bedeuten: Gruppe 0: 0 bzw. 1 Installation
Gruppe 1: 2 bis 5 Installationen
Gruppe 2: 6 bis 10 Installationen
Gruppe 3: 11 bis 20 Installationen
Gruppe 4: 21 bis 50 Installationen
Gruppe 5: mehr als 50 Installationen

Prozentwerte

Tab.2: Installationen von Standardanwendungssoftware (ISIS Software Report 2/1980)

2.2 STANDARDANWENDUNGSSOFTWARE FÜR DIE MARKTFORSCHUNG

Die Lage auf dem Teilmarkt für Standardanwendungssoftware des Bereiches "Marktforschung und Statistik" läßt sich wie folgt beschreiben: "The 'Statistical Software' has germinated and grown without control or management. Little has been known of the number of these programs, their detailed characteristics, and which ones were dangerous." (Francis, S.44)

In einer Untersuchung der International Association for Statistical Computing wurden 1979 111 Programme analysiert und bewertet. Davon waren 80 amerikanischen Ursprungs, die anderen stammten aus Kontinentaleuropa (10), Großbritannien (9), Australien (9) und Kanada (3). Nach Angaben des ISIS-Software Reports (Herausgeber: Nomina Gesellschaft für Wirtschafts- und Verwaltungsregister, München), Ausgabe Januar 1981, ergibt sich auf dem deutschsprachigen Markt (BR Deutschland (D), Österreich (A), Schweiz (CH) und Liechtenstein (FL)) ein Angebot von 15 Programmen, die in Tabelle 3 hinsichtlich Anbieter, Installationen, Verarbeitungsmodus (Stapel/S, Dialog/D), Rechenzentrumsanschluß und Methoden etwas näher analysiert werden sollen:

Analysepunkt	Produkte														
	/ 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
angeboten in:	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
		A	A	A	A	A	A	A	A	A	A	A	A		A
		CH	CH	CH	CH		CH								
seit 19...:	76	68	74	?	74	72	78	80	?	76	?	?	79	?	70
Anbieter															
in D:	1	1		1	1	1		1	1	1	1	1	1	1	5
in A:												1			2
in CH:												1			2
in FL:			1				1								
Installationen															
weltweit:	9	2	4	?	5	81	2	>5	?	500	?	?	0	?	1500
in D, A, CH:	7	1	4	?	5	75	2	>5	?	5	?	?	0	?	?
Verarb.-modus:	SD	S	SD	SD	SD	SD	SD	D	D	D	D	D	SD	SD	SD
RZ-Anschluß:				x					x	x	x	x	x	x	x
Methoden:															
Deskriptive Stat.	x	x				x		x		x	x	x	x	x	x
Tests	x	x				x			x		x				x
Korrelationen								x		x					x
Regression										x					x
Varianzanalyse										x					x
Faktorenanalyse	x									x				x	x
Diskriminanz-A.										x					x
Clusteranalyse	x		x	x	x									x	
Fragebogenüberprüfung															x
Angebotene Programmpakete:															
1	STAR (Survey Tabulation and Reporting System)														
2	PRODASTE (Datenanalyse statistischer Erhebungen)														
3	TYP-2 (Clusteranalyse Typologie)														
4	CLUST (Clusteranalyse)														
5	SEG 2-2 (Clusteranalyse-Segmentation)														
6	RPS (Regionalpass)														
7	NONKORR (Nichtlineare Korrelationsrechnung)														
8	MAFOR 2 (Marktforschungsprogramm-Auswertung)														
9	STATIS (Statistisches Programmpaket)														
10	XCALABER (Frequency Tabulation Program)														
11	SURVEY (Marktforschungsanalyzesystem)														
12	CROSSTABS (Kreuztabellenanalyse)														
13	TAB (Auswertung von Primärdaten)														
14	SPS (Survey Processing System)														
15	SPSS (Statistical Package for Social Sciences)														

Tab.3: Vergleich der auf dem deutschsprachigen Markt angebotenen Standardanwendungssoftware im Bereich marktforschung und Statistik

Von den 15 Programmen wurden zwei (CLUST und TAB) erstmals im ISIS-Report angeführt, weiterhin erhöhten sich die Anbieter von SPSS während des letzten Halbjahres in Deutschland um zwei auf nunmehr fünf.

Alle Pakete sind in Deutschland erhältlich, doch gibt es in den kleineren Märkten in Österreich und der Schweiz eine nur unwesentlich geringere Angebotsdichte.

Die meisten Anbieter haben ihren Sitz in der BR Deutschland, in Österreich treten nur drei Firmen als Anbieter auf, von denen zwei das Programmpaket SPSS vertreiben, das dritte gehört zur Sharp-Gruppe und bietet deren Paket CROSSTABS an.

In der Installationshäufigkeit ergeben sich zwischen den einzelnen Programmen große Differenzen. Wurde ein (allerdings noch relativ junges) Produkt bisher in noch keinem einzigen Unternehmen installiert, so hat es der "Marktrenner" SPSS weltweit schon auf 1500 Implementationen gebracht, die genaue Zahl für den deutschsprachigen Raum ist unbekannt. Die meisten Programme (40%) liegen in Installationsgruppe I (2-5 Installationen), gefolgt von Installationsgruppe V (mehr als 50 Installationen) mit 30% Marktanteil.

Vergleicht man die Struktur dieses Teilmarktes mit der des Gesamtmarktes "Kommerzielle Programme" mittels eines Chi-Quadrat-Tests (Fehlerwahrscheinlichkeit = 5%) auf Gleichheit der Verteilungen, so ergeben sich keine signifikanten Differenzen,

d.h. die Struktur des Teilmarktes für Marktforschung und Statistik kann als durchaus typisch für den Gesamtmarkt angesehen werden.

Die meisten Programme sind sowohl für Dialog- als auch Stapelverarbeitung (SD) geeignet (9), bei fünf Programmen ist ausschließlich Dialog-Betrieb (D) zulässig, und einzig beim ältesten schon seit 1968 auf dem Markt befindlichen Produkt ist nur Stapelverarbeitung (S) möglich.

Etwas mehr als die Hälfte der Programme kann nur gekauft und/oder gemietet werden. Bei den anderen (mit x in der Zeile "Rechenzentrum (RZ)-Anschluß" gekennzeichneten) Produkten bietet der Anschluß an ein DV-Zentrum meist die einzige Möglichkeit, Daten auszuwerten.

Analysiert man die in den Programmen angebotenen Methoden, darunter soll hier die Beschreibung des jeweiligen Lösungsweges verstanden werden, so lassen sich die Pakete in drei Gruppen gliedern:

(1) Produkte, die eine vollständige Analyse von Marktforschungsdaten ermöglichen sollen und daher ein weitreichendes Spektrum an Funktionen beinhalten (STAR, STATIS, SPSS) -

(2) Programme für Spezialanalysen (TYP-2, CLUST und SEG 2-2 zur Clusteranalyse, NONKORR für Korrelationsanalysen und SPS für Faktoren- und Clusteranalyse) - und

(3) Pakete zur Durchführung einfacher Primärdaten-Untersuchungen, das sind deskriptive Statistiken und/oder Tests (PRODASTE, RPS, MAFOR 2, XCALABER ,SURVEY, CROSSTABS und TABS).

Die hier durchgeführte Analyse garantiert keine Vollständigkeit, vor allem in bezug auf die in den Programmen angebotenen Methoden zur Datenanalyse, da sie sich einzig auf die dem ISIS-Report zu entnehmenden Informationen stützt.

Weiterhin führten wir keinen Kostenvergleich durch, da aufgrund der sehr unterschiedlichen Vorgangsweise der Anbieter bei der Preiskalkulation - so entstehen beim Anschluß an ein Rechenzentrum eventuell Gebühren für den Wählleitungsanschluß, den Datentransfer, die Benutzung von CPU und Massenspeicher, die von den verschiedensten Faktoren abhängen und meist der individuellen Vereinbarung unterliegen - keine einigermaßen verlässliche Kostenkalkulation möglich ist.

3. KRITERIEN FÜR DIE AUSWAHL VON STANDARDANWENDUNGSSOFTWARE

In der Literatur gibt es bereits eine große Anzahl an Kriterienkatalogen zur Bewertung von Standardanwendungssoftware.

Wedekind unterscheidet acht Auswahlkriterien (Wedekind, S.152):

1. Systemfunktionen des Anwendungspaketes
2. Systementwurf und Systemimplementation
3. Hardware/Software-Voraussetzungen bei der zur Verfügung stehenden Anlage
4. Dokumentation
5. Reifegrad des Paketes
6. Systembetrieb und Systempflege
7. Erfahrungen und Ruf des Anbieters
8. Kosten (vgl. die ähnliche Kriterienliste von H.R.Hansen (Hansen 1981, S.8)).

Eine andere Unterteilung der Kriterien wählt Frank, der zwischen produkt- und anbieterbezogenen Beurteilungskriterien sowie relevanten Nebenbedingungen unterscheidet (Frank, S.29ff), während Zimmermann bei der Beurteilung von Standardsoftware drei Ebenen betrachtet haben will (Zimmermann, S.300ff, S.419ff, S.504ff):

Primäre Qualitätsmerkmale (Funktionsumfang, Effizienz, Benutzerfreundlichkeit, Flexibilität, Portabilität, Zuverlässigkeit),

Zusatzleistungen und ergänzende Qualitätsmerkmale (Dokumentation, Schulung, Serviceleistungen, Zusatzbausteine, Zukunftssicherung),

und äußere Randbedingungen (Preise, Verfügbarkeit, Konditionen und Restriktionen, Image des Software-Herstellers).

Diese Kriterienkataloge leiden alle an mangelnder Operationalität, da sie zwar umfangreich sind , aber sehr allgemein gehalten werden und zuwenig auf die vom Anwender benötigten Methoden zur Problemlösung Wert legen, d.h. die Kriterienkataloge tragen die Handschrift des Informatikers und nicht die des Betriebswirtes. Einzige Ausnahme bildet ein Katalog zur Beurteilung von Standardanwendungssoftware auf dem Gebiet der Kostenrechnung, der sich auf die in diesem Fachbereich erforderlichen Verfahren beschränkt (vgl. Hansen 1979, S.152).

Bei der bereits oben erwähnten US-Studie zur Beurteilung von Standardanwendungssoftware des Bereiches "Marktforschung und Statistik" wurde ein primär anwenderbezogener Katalog ("Four questions that a potential user would have") mit folgenden Schwerpunkten benutzt (Francis, S.45):

1.Capabilities

1.1.Processing and displaying data

1.1.1.File building and manipulation

1.1.2.Editing

1.1.3.Data Display

1.2.Exploration and mathematical analysis of statistical data

2.Portability

3.Ease of learning and using

4.Reliability

Im Gegensatz dazu ist der hier erstellte Kriterienkatalog problemorientiert, ohne die Wichtigkeit von Dateneingabe und -manipulation sowie die Gestaltung der Ausgabe zu vernachlässigen. Abschnitt 1 und die Exkurse 1 und 2 befassen sich mit Methoden, die zur Auswertung des bei Marktforschungsuntersuchungen gewonnenen Datenmaterials notwendig sind, die Abschnitte 2 und 3 beurteilen Daten- und Datei-Handling und Ausgabegestaltung:

1.1 Häufigkeitsauswertungen

2 Tests

3 Korrelationen

4 Regression

5 Varianzanalyse

6 Faktorenanalyse

7 Diskriminanzanalyse

8 Clusteranalyse

2 Daten- und Datei-Handling

3 Ausgabegestaltung

Exkurs 1: Überprüfung von Fragebogenitems

Exkurs 2: Zeitreihenanalyse

Aus Erfahrungen der beiden Autoren bei der empirischen Durchführung von Marktforschungsstudien läßt sich ableiten, daß die mit Hilfe von Häufigkeitsauswertungen und Tests gewonnenen Ergebnisse in der Praxis die größte Rolle spielen. Die Anwendung der anderen Methoden scheitert i.d.R. an der mangelnden Qualität der Daten (beispielsweise bezüglich des Skalierungsniveaus)

oder am mangelnden Know-How des entscheidungsbefugten Managements. Daher wurde bei der Erstellung des Detail-Kriterienkataloges in den beiden gebräuchlichsten Methodenbereichen darauf geachtet, daß bei der Untergliederung eine möglichst große Vollständigkeit durch Aufzählen aller Kriterien erreicht wird, während dies bei Verfahren teilweise unterblieb (beispielsweise beim Kriterium "Statistische Kennzahlen").

Die Überprüfung von Fragebogenitems ist nur als Exkurs im Katalog enthalten, da dieses Gebiet als Randbereich der Auswertung angesehen werden kann. Die Kriterien dieses Kapitels sollten daher bei der Auswahl von Standardanwendungssoftware bestenfalls bedingt herangezogen werden, obwohl die potenten Marktforschungspakete (zB SPSS und das im ISIS-Report nicht enthaltene SAS) auch für diesen Funktionsbereich Methoden beinhalten.

Ähnlich verhält es sich bei der Zeitreihenanalyse. Das Datenmaterial für Marktforschungsuntersuchungen stammt zumeist nicht aus Zeitreihenuntersuchungen, trotzdem bieten die verschiedenen Programme ein (sehr unterschiedlich breites) Spektrum an Prozeduren zur Analyse an. Deshalb ist die Zeitreihenanalyse als Exkurs im Übersichtskatalog enthalten, ohne auf sie im Detail-Kriterienkatalog noch näher einzugehen. Dieser Schritt geschieht in Übereinstimmung mit dem ISIS-Software Report, der ebenfalls zwischen den Bereichen "Zeitreihenanalyse" und "Marktforschung" unterscheidet.

Innerhalb der Kapitel des Kriterienkatalogs erfolgt eine
erfolgt eine Dreiteilung in

- Eingabe,
- Verarbeitung und
- Ausgabe.

Wir hoffen durch diese Trennung eine bessere Übersicht über
die Kriterien eines Kapitels und eine klarere Abgrenzung zwi-
schen ihnen zu erreichen (vgl. dazu auch C.Jordan).

In den Sternchen-Raster kann im Zuge einer Nutzwertanalyse
die maximale Punktezahl je Beurteilungsmerkmal eingetragen
werden, wodurch eine Gewichtung der Methodenbereiche und
der Kapitel Daten- und Datei-Handling und Ausgabegestaltung,
aber auch der einzelnen Kriterien erfolgt. Innerhalb des
Punkterahmens können in weiterer Folge alternative Programm-
pakete bewertet werden (vgl. Hansen / Bauer).

4. KRITERIENÜBERSICHT

4.1 METHODEN

1. HAUFIGKEITSAUSWERTUNGEN	**	**
Eingabe:	**	**
1. Daten	*	*
2. Variablenfestlegung	*	*
3. Methodenaufruf	*	*
Verarbeitung:	**	**
1. Deskriptive statistische Kennzahlen für eine Variable	*	*
2. Deskriptive Statistiken für Fallgruppen	*	*
3. Mehrdimensionale Häufigkeitsverteilungen	*	*
4. Häufigkeitstabellen für Variablen mit Mehrfachantworten	*	*
Ausgabe:	**	**
1. Häufigkeitswerte	*	*
2. Statistiktabelle	*	*
2. TESTS	**	**
Eingabe:	**	**
1. Daten	*	*
2. Klassifizierungsvariablen	*	*
3. Testvariablen	*	*
4. Methodenaufruf	*	*
Verarbeitung:	**	**
1. Parametrische Tests	*	*
2. Nichtparametrische Tests	*	*
Ausgabe:	**	**
1. Testergebnisse	*	*
2. Teststatistik	*	*
3. KORRELATIONEN	**	**
Eingabe:	**	**
1. Daten	*	*
2. Variablenfestlegung	*	*
3. Bestimmung der Korrelationsbeziehungen	*	*
4. Methodenaufruf	*	*

Verarbeitung:	**	**
1. Pearson's Produkt-Moment-Korrelationskoeffizient	*	*
2. Partieller Korrelationskoeffizient	*	*
3. Kanonische Korrelation	*	*
4. Spearman's Korrelationskoeffizient	*	*
5. Kendall's Korrelationskoeffizient	*	*
Ausgabe:	**	**
1. Tabelle der Korrelationskoeffizienten	*	*
2. Statistische Maßzahlen	*	*
 4. REGRESSION	 **	 **
Eingabe:	**	**
1. Daten	*	*
2. Variablenfestlegung	*	*
3. Methodenaufruf	*	*
4. Verfahrenssteuerung	*	*
Verarbeitung:	**	**
1. Einfachregression	*	*
2. Multiple Regression	*	*
3. Nichtlineare Regression	*	*
4. Autoregression	*	*
5. Regression in einem System linearer interdependenter Gleichungen	*	*
6. Sonstige Regressionsverfahren	*	*
7. Statistische Kennzahlen	*	*
Ausgabe:	**	**
1. Regressions- und Statistiktabelle	*	*
2. Varianzanalyse	*	*
 5. VARIANZANALYSE	 **	 **
Eingabe:	**	**
1. Daten	*	*
2. Methodenaufruf	*	*
3. Variablenfestlegung	*	*
4. Testaufruf	*	*
Verarbeitung:	**	**
1. Einfaktorielle Varianzanalyse	*	*
2. Mehrfaktorielle Varianzanalyse	*	*
3. Kovarianzanalyse	*	*
4. Berechnung der Effekte	*	*
5. Regressionskoeffizienten	*	*
6. Tests	*	*
7. Gruppenstatistiken	*	*

Ausgabe:		**	**
1. Varianzanalysetafel		*	*
2. Testergebnisse		*	*
3. Statistiken		*	*
4. Matrizen		*	*
6. FAKTORENANALYSE		**	**
Eingabe:		**	**
1. Daten		*	*
2. Methodenaufruf		*	*
3. Faktorenfestlegung		*	*
4. Kommunalitäten		*	*
Verarbeitung:		**	**
1. Datenreduktion		*	*
2. Rotation		*	*
3. Berechnung der Korrelationsmatrix		*	*
4. Varianzzerlegung		*	*
5. Faktorenmatrizen		*	*
Ausgabe:		**	**
1. Korrelationsmatrix		*	*
2. Faktorenwerte		*	*
3. Faktormuster		*	*
4. Matrizen		*	*
7. DISKRIMINANZANALYSE		**	**
Eingabe:		**	**
1. Daten		*	*
2. Methodenaufruf		*	*
3. Variablenfestlegung		*	*
Verarbeitung:		**	**
1. Analysephase		*	*
2. Berechnungen		*	*
- Kovarianzen			
- Korrelationen			
- Diskriminanzfunktionen			
- Klassifikationsfunktionen			
3. Klassifikationsphase		*	*
Ausgabe:		**	**
1. Diskriminanzwerte		*	*
2. Klassifikationsergebnisse		*	*
3. Territoriumskarte		*	*
4. Matrizen		*	*

8. CLUSTERANALYSE

Eingabe:

1. Daten
2. Methodenaufruf

Verarbeitung:

1. Vorgangsweise
 - nicht hierarchisch
 - hierarchisch
2. Berechnungen
 - Standardisierung
 - Distanzmaße
 - Statistiken der Variablen
 - Statistiken der Cluster
 - Tests

Ausgabe:

1. Protokoll
2. Dendrogramm

4.2 DATEN- UND DATEI-HANDLING

Eingabe:

1. Variable
2. Etiketten
3. Formate
4. Fehlende Werte
5. Dateinamen
6. Zusatztext
7. Eingriff in die Speicherverwaltung

Aufbereitung von Dateien:

1. Sortieren
2. Einbeziehungen
3. Änderungen
4. Berechnungen
5. Hinzufügungen
6. Löschen
7. Zusammenfassungen

Arbeiten mit Dateien:

1. Selektion
2. Bedingungen und Schleifen
3. Heranziehungsmöglichkeiten
4. Multiplikationen
5. Standardisierung
6. Datenbankanschluß
7. Kopieren
8. Dokumentation

4.3 AUSGABEGESTALTUNG

1. Ausgabeinheiten
2. Charts
3. Report Writing

** **

** **

* *

* *

** **

* *

* *

** **

* *

* *

** **

** **

* *

* *

* *

* *

* *

* *

* *

** **

* *

* *

* *

* *

* *

* *

* *

** **

* *

* *

* *

* *

* *

* *

* *

* *

** **

** **

** **

** **

Exkurs 1:

ÜBERPRÜFUNG VON FRAGEBOGENITEMS

** **

Eingabe:

** **

1. Daten * *
2. Bestimmung der Skalen * *
3. Festlegung der Variablen, deren Zusammenhang
Überprüft werden soll * *
4. Festlegung der Trennstellen * *
5. Methodenaufruf * *

Verarbeitung:

** **

1. Dichotomisierung der Variablen * *
2. Guttman-Skala * *
3. Statistische Kennzahlen für Guttman-Skala * *
4. Rehabilitätskoeffizienten * *
5. Statistische Maßzahlen der Items * *
6. Tests für das Verhältnis der Items zueinander * *

Ausgabe:

** **

1. Skalogramm für Guttman-Analyse * *
2. Rehabilitätstabelle * *
3. Varianzanalyse * *
4. Testergebnisse * *

Exkurs 2:

ZEITREIHENANALYSE

Eingabe:

1. Daten
2. Methodenaufruf
 - Angabe eines Trendverlaufs der Zeitreihe
 - Anzahl der vorauszusagenden Perioden
 - Gewichtung

Verarbeitung:

1. Univariate Prognosemethoden
 - horizontales Modell
 - lineares Modell
 - quadratisches Modell
2. Autoregressives Modell
3. Box-Jenkins-Modell (ARIMA)
4. Adaptive Filterung
5. Saisonmodelle
6. Multivariate Prognosemethoden
7. Verfahren von Matt
8. Spektralanalyse

Ausgabe:

1. Vorhersagewerte
 - schriftlich
 - grafisch
2. Statistiktabelle

5. ERLÄUTERUNG DER METHODEN

In den nachfolgenden Ausführungen sollen die im Kriterienkatalog aufscheinenden Methoden kurz erläutert werden.

5.1 HÄUFIGKEITSAUSWERTUNGEN

Die einfachste und am meisten verbreitete Auswertung von Marktforschungsdaten erfolgt durch Ermittlung der Häufigkeitsverteilungen. Hierbei sind vier alternative Vorgangsweisen zu unterscheiden:

Bei der Ermittlung deskriptiver statistischer Kennzahlen für eine Variable, wobei in den meisten Fällen zwischen diskreten und stetigen Variablen unterschieden wird, soll diese auf ihre statistischen Merkmale überprüft werden. Die für eine Marktforschungsuntersuchung relevanten Kennzahlen sind im Detail-Kriterienkatalog aufgelistet.

Eine Erweiterung gegenüber diesen einfachen Statistiken stellen deskriptive Statistiken für Fallgruppen dar. Nicht mehr für den gesamten Bereich einer Variablen werden Kennzahlen berechnet (z.B. Einwohner der Stadt Wien), sondern gesondert für Teile derselben (z.B. für männliche und weibliche Bewohner getrennt).

Etwas anspruchsvoller als die oben beschriebenen Prozeduren ist die Erstellung mehrdimensionaler Häufigkeitsverteilungen. Hiermit können die Häufigkeitsverteilungen von Variablen nicht nur nach einer, sondern nach n-Dimensionen untergliedert werden, was zur Berechnung von zwei- bis n-dimensionalen Kontingenztafeln (Kreuztabellen) führt.

Besondere Bedeutung hat die Auswertung von Variablen mit Mehrfachantworten. Da diese in Marktforschungsuntersuchungen sehr häufig vorkommen (z.B. auf die Frage: Welche Sportarten betreiben Sie ?) sind Prozeduren zur Analyse der Häufigkeitsverteilungen sowie schnellen und einfachen Erstellung von relevanten Kreuztabellierungen in einem Standardanwendungssoftware-Paket für die Marktforschung von großer Wichtigkeit.

5.2 TESTS

Die für die Marktforschung relevanten Testmethoden "befassen sich mit der Übertragung der aus der Stichprobe gewonnenen und zunächst nur für diese gültigen Ergebnisse auf die Grundgesamtheit" (Berekoven, S.172) und mit der Ermittlung von signifikanten Unterschieden zwischen Teilen der Stichprobe.

Es lassen sich zwei große Gruppen von Testverfahren unterscheiden, die im folgenden erläutert werden sollen.

Die bekanntesten der in der Marktforschung gebräuchlichen Tests fallen in die Klasse der parametrischen Testverfahren. Um die-

se anwenden zu können, müssen die dem jeweiligen Test zugrundeliegenden Daten bestimmten Anforderungen in bezug auf ihre Verteilung und Skalierung genügen. Dies führt oft zu Problemen bei der Durchführung von Tests und schränkt die Aussagekraft von durch Tests gewonnenen Ergebnissen erheblich ein.

Anwendung finden parametrische Testverfahren vor allem bei Vergleichen zwischen zwei bzw. n Mittelwerten und um Aussagen über signifikante Unterschiede zwischen den anderen Parametern der empirisch ermittelten Verteilungen zu erhalten.

Die Gruppe der nichtparametrischen Testverfahren ist in der Praxis noch nicht sehr weit verbreitet. Mit diesen Methoden können Aussagen über signifikante Verteilungsunterschiede getroffen werden, ohne irgendwelchen Beschränkungen in der Verteilung der zu testenden Daten zu unterliegen. Weiterhin können diese Tests teilweise auch bei ordinal- und nominalskalierten Stichprobenwerten angewandt werden, während für parametrische Tests immer metrisch bzw. intervallskalierte Daten vorliegen müssen.

5.3 KORRELATIONEN

In den Korrelationsverfahren wird der statistische Zusammenhang zwischen zwei oder mehreren Variablen berechnet. Die diversen Korrelationskoeffizienten, die in diesem Kapitel näher erläutert werden sollen, können Werte zwischen 0 (kein Zusammenhang) und ± 1 (perfekter Zusammenhang) annehmen.

Der Pearson Produkt-Moment-Korrelationskoeffizient mißt den Zusammenhang zwischen zwei Variablen, die zumindest intervallskaliert sein müssen. Das Quadrat dieses Koeffizienten, der in der Literatur von allen Korrelationskoeffizienten den höchsten Bekanntheitsgrad hat, heißt das Bestimmtheitsmaß und gibt an, welcher Teil der Varianz einer Variablen durch die andere Variable erklärt wird, wobei man eine lineare Regressionsfunktion zugrunde legt.

Der partielle Korrelationskoeffizient ermittelt die Korrelation zwischen zwei Variablen unter Ausschaltung des Einflusses von vorher bestimmten Kontrollvariablen, d.h. es wird quasi die "reine Korrelation" zweier Variablen errechnet, allerdings wiederum unter den Bedingungen einer linearen Regressionsfunktion und intervallskalierten Daten.

Die kanonische Korrelation stellt eine Erweiterung der multiplen Regression dar. Statt des Zusammenhangs eines Variablensatzes mit einer Variablen (wie bei der multiplen Regression) wird die Korrelation zwischen zwei Variablensätzen festgestellt.

Die Rangkorrelationskoeffizienten von Spearman und Kendall setzen weder Normalverteilung in der Stichprobe noch intervallskalierte Daten voraus, so daß ihnen in Marktforschungsuntersuchungen, in denen oftmals ordinalskaliertes Datenmaterial überwiegt, große Bedeutung zukommt.

Spearman's Koeffizient ist bei annähernd kontinuierlichen Daten, das bedeutet, daß nur wenige Werte denselben Rang innehaben,

vorzuziehen, während Kendall's Koeffizient bei einer großen Anzahl von Beobachtungen, die wenigen Rängen mit folglich hoher Besetzungszahl zugeordnet werden, bessere Ergebnisse liefert.

5.4 REGRESSION

Aufgabe der Regressionsanalyse ist die Erfassung von Abhängigkeiten zwischen Variablen mit Intervall- oder Verhältnisskalenniveau. Es gilt diejenige Funktion zu finden, die diesen Zusammenhang am besten beschreibt. Im folgenden möchten wir diejenigen Regressionsverfahren, die sich im Kriterienkatalog finden, kurz charakterisieren.

Die Einfachregression stellt das Grundmodell der Regressionsanalyse dar. Dieses Verfahren behandelt den linearen Zusammenhang zwischen nur zwei Variablen, wovon die eine (= unabhängige) zur Erklärung der anderen (= abhängigen) herangezogen wird.

Die multiple Regression erweitert das Grundmodell. Anstelle einer einzigen unabhängigen Variablen wird ein ganzer Variablenansatz benutzt, um eine abhängige Variable zu erklären - allerdings bleibt die Annahme eines linearen Funktionsverlaufs aufrecht.

Diese Bedingung wird bei der nichtlinearen Regression aufgegeben. Damit ergibt sich für das Verfahren ein erweiterter

Anwendungsbereich, jedoch muß die Erhöhung der Komplexität der Lösungssuche als Einschränkung betrachtet werden.

Die Autoregression dient primär zur Zeitreihenanalyse. Aufgrund einer (zeitlichen) Verschiebung der Daten innerhalb der Zeitreihe können Regelmäßigkeiten in derselben erkannt, Zusammenhänge zwischen Daten aus verschiedenen Perioden berechnet und in der Folge saisonale und auf Trends beruhende Veränderungen des Datenmaterials ermittelt werden.

In Modellen, die aus einem System linearer, voneinander abhängiger Gleichungen bestehen, bereitet die Lösungsfindung aufgrund der Interdependenz zwischen abhängigen und unabhängigen Variablen der diversen Gleichungen große Probleme. Mit Hilfe eines Sets spezieller Regressionsverfahren, von denen jeweils das effizienteste ausgewählt wird, soll diese Aufgabenstellung, die sehr viel Rechenzeitaufwand erfordert, gelöst werden.

Von weiteren Regressionsverfahren, die bei Marktforschungsstudien eine Rolle spielen, ist die Regression mit Dummy-Variablen die wichtigste. Hiermit können die in diesem Bereich relativ oft auftretenden qualitativen Variablen durch Einführung von Variablen, die nur die Werte 0 und 1 annehmen können, "quantifiziert" werden.

Erwähnenswert sind noch die gewichtete Regression, mit der eine unterschiedliche Gewichtung der Abstände zwischen Voraussage- und Beobachtungswert und somit eine Systemsteuerung möglich ist, sowie ein Verfahren zur Ermittlung der Regres-

sionskoeffizienten für jede Kombination zwischen vorher bestimmten abhängigen und unabhängigen Variablen. Die Anwendung dieses Verfahrens ist allerdings nur bei Untersuchungen einer größeren Anzahl von Modellen sinnvoll.

5.5 VARIANZANALYSE

In der von R.A. Fisher entwickelten Varianzanalyse benötigt man eine zumindestens intervallskalierte abhängige Variable (sog. Kriteriumsvariable). Falls die unabhängigen Variablen zum Teil nicht-metrischer Skalierung sind, wird von Faktoren gesprochen.

Interessiert man sich für die möglichen Effekte eines Faktors, liegt eine eindimensionale Varianzanalyse vor. Werden die simultanen Effekte von mehreren (insgesamt n) Faktoren betrachtet, erfolgt eine n -dimensionale Varianzanalyse. Die Grundidee der Varianzanalyse ist die Zerlegung der Gesamtvarianz in zwei Komponenten. Die erste ist jener Teil der Varianz, der auf den Faktor (die Faktoren) zurückzuführen ist, sich also auf die Varianz zwischen den durch den (die) Faktor(en) erzeugten Kategorien bezieht. Die zweite Komponente steht für die Varianz innerhalb der Kategorien.

Daher ist die Heranziehung der Varianzanalyse beispielsweise zur Erklärung von Einkommensstreuungen bzw. Budget-

angaben sinnvoll, wenn nominalskalierte (z.B. Geschlecht) oder ordinalskalierte Variable (z.B. Altersklassen) eine Unterteilung der Stichprobe in Sub-Sets ermöglichen.

Das klassische Experimentdesign teilt die Fehlerquadratsumme in Anteile, die erstens auf additive Effekte der Faktoren (z.B. Faktoren A und B), zweitens auf Interaktionseffekte zwischen A und B, sowie drittens auf Fehler (Zufallsabweichungen) zurückzuführen sind. Mittels F-Test können nun die Signifikanzniveaus des Gesamtmodells, der Interaktion, des additiven Modells und jedes Einzeleffekts überprüft werden, wobei die jeweilige Varianzkomponente zur Restkomponente in Beziehung gesetzt wird.

Der Hauptunterschied zum orthogonalen Modell für gleiche Besetzungszahlen in den Kategorien liegt darin, daß bei starken Beziehungen zwischen den Faktoren A und B der additive Effekt signifikant sein kann, auch wenn dies die Einzeleffekte nicht sind.

Das hierarchische Modell testet die Signifikanz stufenweise, sodaß die Reihenfolge der Einbeziehung der Faktoren entsprechend einer subjektiven Voreinschätzung der Bedeutung derselben angegeben werden kann, was auch von Wichtigkeit ist, wenn man zu einzelnen Faktoren bzw. den diesbezüglichen Angaben weniger Vertrauen hat.

Der klassische Regressionsansatz überprüft einen Effekt unter Bereinigung aller anderen Effekte (auch der Interaktionen), sodaß der "Netto-Einfluß" deutlich wird.

Von Ansätzen der Kovarianzanalyse wird gesprochen, wenn metrische unabhängige Variable (Kovariate) mit nicht-metrischen Faktoren in einem Untersuchungsdesign kombiniert werden, wobei diese Kovariate die Streuung in den Angaben zur abhängigen Variable erklären sollen, wonach eine konventionelle Varianzanalyse für die "korrigierten" Werte durchgeführt wird.

Mittels Tests kann nicht nur die Varianzhomogenität sondern auch die Trennschärfe zwischen den Gruppen überprüft werden.

Hilfreiche Gruppenstatistiken sind die Anzahl der Fälle für jede Gruppe, Mittelwert, Standardabweichung, Minimum, Maximum und Konfidenzintervall für den Mittelwert der Variable, deren Streuung erklärt werden soll. Weiters können die Effekte in feste und zufällige geteilt werden. Der Eta-Quadrat-Wert gibt den durch den Faktor erklärten Varianzanteil an. Zusammenhänge zeigen sich aus dem Beta-Koeffizient der partiellen Korrelation bzw. aus den Korrelationswerten zwischen der abhängigen Variablen und den Faktoren, Kovariaten und Faktor-Faktor-Interaktionen.

5.6 FAKTORENANALYSE

Ausgehend von einer Vielzahl von (korrelierten) Variablen trägt die Faktorenanalyse hauptsächlich dafür Sorge, daß wenige, noch zu interpretierende Faktoren oder Komponenten ermittelt werden, die nun als "Quellvariable" für die Interrelationen der Daten gelten. Neben dieser Erkennungsmöglichkeit von Mustern und der wichtigen Datenreduktionsfunktion (eigentlich Variablenreduktion) kann die Faktorenanalyse außerdem zum Abtesten von Hypothesen über mittels der Faktoren beschriebene Variable und zur Konstruktion von Indizes, die als neue Variablen in spätere Analysen eingehen, benützt werden. Je nachdem, ob die herangezogene Korrelationsmatrix den Zusammenhang zwischen Objekten, Individuen etc. oder aber zwischen Variablen ausdrückt, kann zwischen einer Q-Faktorenanalyse und einer R-Faktorenanalyse unterschieden werden.

Die Hauptkomponentenmethode gilt als rechenzeitparend, da eine Invertierung der Korrelationsmatrix nicht nötig ist. Diese Version kann selbst noch bei hochkorrelierten Variablen eingesetzt werden. Die ermittelten Eigenwerte in der Hauptkomponenten-Matrix geben den Varianzanteil an, den der jeweilige Faktor bestimmt. Eine Ausweitung dieser Methode ermöglicht Iterationen zur Verbesserung der Kommunalitäten-Schätzwerte.

Rao's Canonical Factoring sucht jene Faktorenlösung, in der die Korrelation zwischen dem Satz der hypothetischen Faktoren und dem Satz der Datenvariablen maximal ist. Die Version Rao betrachtet die Daten als Stichprobe und schließt auf eine (hypothetische) Grundgesamtheit.

Eine Alpha-Faktorenanalyse geht ebenfalls nach dem Stichprobenkonzept vor, sucht aber Faktoren maximaler Generalisierbarkeit und zieht Kommunalitäten statt Varianzen zur Skalierung heran.

Die Image-Faktorenanalyse nimmt an, daß das multiple Bestimmtheitsmaß den besten Schätzwert für den allgemein beeinflussten Teil der Variable darstellt, was in der standardisierten Form als Image bezeichnet wird. Die Ergänzung auf den vollständigen standardisierten Variablenwert heißt Anti-Image und schätzt den "unique"-Anteil in der Variable, der möglichst klein sein sollte (vgl. z.B. Nie et al., S.479ff).

An die Phase der Datenreduktion schließt sich die Rotation, wobei die einzelnen Versionen sich in der Art der Umformung der Faktorenmatrix unterscheiden. Durch die Rotation soll eine eindeutige Zuordnung der Variablen zu Faktoren erzielt bzw. die Interpretationsmöglichkeit erhöht werden. Die Zusammenhänge zwischen den vorhergehenden und den rotierten Faktorenwerten zeigen sich der Transformationsmatrix. Die aus den Interfaktorenkorrelationen ersichtlichen Zusammenhänge zwischen den Faktoren

sollten möglichst gering sein. Die Faktorenwertkoeffizienten können zur Datenreduzierung für neue Fälle herangezogen werden.

5.7 DISKRIMINANZANALYSE

Die Diskriminanzanalyse geht dem Wunsch nach, zwischen 2 oder mehr Gruppen, die sich in einer speziellen Situation ergeben (bzw. wenn die Zuordnung zu denselben aufgrund erhobener Merkmale überprüfbar ist), statistisch zu unterscheiden. Dafür werden diskriminierende Variable ausgewählt und auf ihre Diskriminierungsstärke abgetestet, welche die Charakteristika, in denen die Gruppen Unterschiede aufweisen, aufzeigen sollen.

Der Analyseaspekt der Diskriminanzanalyse liegt in der Offenlegung der Stärke der diskriminierenden Variablen und in der eventuellen Reduzierung der Anzahl der Diskriminanzfunktionen. Die schrittweise Einbeziehung von Diskriminanzvariablen (die wichtigsten zuerst) hilft bei einer zu großen Zahl von zur Auswahl stehenden Variablen.

Liegt eine zufriedenstellende Diskriminanzfunktion vor, können noch nicht einbezogene Elemente (Fälle) damit klassifiziert werden, sodaß es unter Umständen möglich ist, nur an Hand von äußeren Merkmalen (z.B. Geschlecht, Altersgruppe etc.) die einzelnen Personen der jeweiligen "wahrschein-

lichsten" Gruppe zuzuordnen. Zur Überprüfung der Adäquanz empfiehlt sich allerdings zuvor eine Test-Klassifikation der Originaldaten.

Aus dem Eigenwert ist die relative Bedeutung der jeweiligen Diskriminanzfunktion ersichtlich. Eine weitere Beurteilungshilfe stellt die kanonische Korrelation dar.

Die Klassifikationsgleichungen werden von der zusammengefaßten Gruppen-Kovarianz-Matrix und den Zentroiden der Diskriminanzvariablen abgeleitet.

Der einzelne Fall wird nun jener Gruppe zugeordnet, für deren Klassifikationsfunktion er den höchsten Wert erzielt. Unter der Annahme einer multivariaten Normalverteilung können die Klassifikationswerte in Wahrscheinlichkeiten der Gruppenzugehörigkeit umgewandelt werden. An Hand der Trefferwahrscheinlichkeiten bzw. des Prozentsatzes der richtigen Klassifizierungen der Originaldaten zeigt sich die Qualität der Trennung zwischen den Gruppen.

5.8 CLUSTERANALYSE

Clusteranalysen haben die Funktion der Datenstrukturierung. Als Ablaufschema zur Durchführung einer Clusteranalyse führen Steinhausen/Langer (S.19ff) an:

1. Präzisierung der Untersuchungsfrage
2. Auswahl der Elemente bzw. Variablen
3. Aufbereitung der Daten (Rohdatenmatrix)
4. Festlegung der Ähnlichkeits- bzw. Distanzfunktion
5. Bestimmung des Algorithmus
6. Technische Durchführung
7. Analyse der Ergebnisse
8. Interpretation der Ergebnisse

Bei einer hierarchischen Vorgangsweise wird die ursprünglich große Anzahl von Clustern (im Extremfall der Fallzahl entsprechend) durch Zusammenlegung schrittweise reduziert. Dadurch werden die zu untersuchenden Alternativen im weiteren Verlauf der Analyse verringert. Eine fehlerhafte Agglomeration bleibt aber bis ans Ende erhalten, da eine vollzogene Zusammenfassung in einem späteren Schritt nicht mehr rückgängig gemacht wird. Die hierarchischen Methoden sind entweder Linkage-Verfahren oder orientieren sich am Cluster-Mittelwert-Vektor oder an der Fehlerquadratsumme.

Nicht-hierarchische Cluster-Methoden gehen von einer a-priori-Zuteilung der Elemente zu Clustern aus, die verbessert werden

soll. Auch das Ergebnis einer anderen Clusteranalyse kann als Startkonfiguration herangezogen werden.

Neben einem genauen Vorgangsprotokoll und einem graphischen Output über die Zusammenfassungen von Clustern dienen vor allem Mittelwerte und Varianzen der Ausprägungen in den einzelnen Clustern der Interpretation dieser Sub-Sets. In diesem Zusammenhang sind vor allem t-Werte und F-Werte hervorzuheben, die die Signifikanz von Differenzen zwischen den Clustern beurteilen. Die für den jeweiligen Cluster bedeutendsten Variablen, die für die Interpretation desselben herangezogen werden können, weisen sich durch geringe Streuung im Cluster im Verhältnis zur Streuung in der gesamten Objektmenge aus.

Mit Hilfe der Clusteranalyse kann somit eine Zusammenfassung von Elementen erfolgen, wie sie etwa im Zusammenhang mit einer Marktsegmentierung von Bedeutung ist. Die Statistiken zu den Cluster-Variablen ermöglichen dabei die notwendige Charakterisierung der Cluster und damit die Ableitung von Hinweisen für eine differenzierte Ansprache der einzelnen Zielgruppen.

EXKURS 1: ÜBERPRÜFUNG VON FRAGEBOGENITEMS

Die Überprüfung eines Fragebogens kann sich auf zwei Bereiche erstrecken:

Erstens wird mittels Skalenanalyse geprüft, ob die Testitems zu einer eindimensionalen Skala gehören, d.h. ob sie sich lediglich im Schwierigkeitsgrad unterscheiden. Die Auswertung erfolgt durch eine Guttman-Skala, die zeigt, daß bei Lösung eines Items der Schwierigkeit a (definiert als der Prozentsatz von Fällen, die dieses Item richtig gelöst haben) auch alle Items mit geringerem Schwierigkeitsgrad gelöst werden und vice versa.

Zweitens können durch Testhalbierung und Berechnung der inneren Konsistenz eines Fragebogens (sog. Itemanalyse) Zusammenhänge zwischen einzelnen Angaben oder Fragebogenitems sowie die Güte der Testergebnisse ermittelt werden. Dies wird durch Berechnung der Reliabilitätskoeffizienten und das Testen des Verhältnisses der Items zueinander (z.B. Test der Itemmittelwerte und Varianzanalysen) erreicht.

6. D E T A I L - K R I T E R I E N K A T A L O G

6.1 METHODEN

6.1.1 HAUFIGKEITSAUSWERTUNGEN

Eingabe:

1. Daten
 - maximaler Wert numerischer Daten
 - ganzzahlige Werte
 - Dezimalzahlen
2. Variablenfestlegung
 - maximale Anzahl an Variablen
 - maximale Anzahl an Schlüsselwörtern
 - Beschränkungen hinsichtlich alphanumerischer Variablen
3. Methodenaufruf
 - Behandlung von fehlenden Werten
 - Nichtberücksichtigung
 - Einbeziehung in die Berechnung
 - Datenverkürzung
 - Auf- und Abrunden
 - Abschneiden
 - Wertebereich für Berechnungen
 - Gewichtung von Beobachtungen
 - Beschränkungen bei Fallgruppen
 - maximale Anzahl an unabhängigen Variablen
 - maximale Anzahl an Fallgruppen
 - Wertebereich für die Variablen
 - Beschränkungen bei mehrdimensionalen Häufigkeitsverteilungen
 - maximale Anzahl an Dimensionen
 - maximale Anzahl an Variablen
 - maximale Anzahl an Variablenausprägungen
 - maximale Anzahl an Tabellendefinitionen
 - Beschränkungen bei der Auswertung von Mehrfachantworten
 - maximale Anzahl an Gruppennamen
 - maximale Anzahl an Tabellengruppen
 - maximale Anzahl an Dimensionen
 - maximale Anzahl an Mehrfachantwortgruppen

Verarbeitung:

1. Deskriptive statistische Kennzahlen für eine Variable
 - Variable mit stetiger Verteilung
 - Mittelwert
 - Standardabweichung
 - Varianz
 - Variationskoeffizient
 - Standardfehler des Mittelwertes
 - Konfidenzintervall für den Mittelwert
 - Schiefe
 - Wölbung
 - Minimum
 - Maximum

- Range
- Summe
- Transformation auf Mittelwert 0
- Variable mit diskreter Verteilung
 - Häufigkeit der einzelnen Beobachtungen
 - absolut
 - prozentuell
 - kumuliert absolut und in Prozent
 - bereinigt und unbereinigt von fehlenden Werten
 - Mittelwert
 - Standardabweichung
 - Varianz
 - Variationskoeffizient
 - Standardfehler des Mittelwertes
 - Konfidenzintervall für den Mittelwert
 - Median
 - Modus
 - Schiefe
 - Wölbung
 - Minimum und beliebige Anzahl der kleinsten Werte
 - Maximum und beliebige Anzahl der größten Werte
 - Quantile
 - Percentile
 - Range
 - Summe
- 2. Deskriptive Statistiken für Fallgruppen
 - Statistische Kennzahlen innerhalb einer Gruppe
 - Mittelwert
 - Standardabweichung
 - Varianz
 - Summe
 - Statistische Kennzahlen für Beziehung zwischen den Gruppen
 - Korrelationskoeffizienten
 - Varianzanalyse
- 3. Mehrdimensionale Häufigkeitsverteilungen (Kontingenztafeln)
 - Zellenhäufigkeiten
 - absolut
 - prozentuell
 - Zeilenanteil
 - Spaltenanteil
 - Anteil an der Gesamtsumme
 - Unabhängigkeitstests
 - erwartete Zellenhäufigkeit bei Unabhängigkeit
 - Abweichung zwischen beobachteter und erwarteter Zellenhäufigkeit
 - Chi-Quadrat-Wert pro Zelle
 - Statistische Kennzahlen
 - Kontingenzkoeffizient
 - Korrelationskoeffizienten
- 4. Häufigkeitstabellen für Variablen mit Mehrfachantworten
 - Berechnungsart
 - dichotomisierte Variablen
 - aufzählende Variable

- Zellenhäufigkeiten
 - absolut
 - prozentuell
 - in % der Gesamtzahl an gegebenen Antworten
 - in % der gültigen Fälle
 - in % der Zeilensumme
 - in % der Spaltensumme

Ausgabe:

1. Häufigkeitswerte
 - tabellarisch
 - seriell
 - grafisch
 - Baumdiagramm (bei Fallgruppen)
2. Statistiktabelle

6.1.2 TESTS

Eingabe:

1. Daten
 - maximale Anzahl an Beobachtungen
 - Skalierungsniveau
 - metrisch
 - ordinal
 - nominal
2. Klassifizierungsvariablen
 - Variablen
 - numerisch
 - alphanumerisch
 - maximale Anzahl an Klassifizierungsvariablen
3. Testvariablen
 - Variablen
 - numerisch
 - alphanumerisch
 - maximale Anzahl an Testvariablen
4. Methodenaufruf
 - Behandlung von fehlenden Werten
 - Festlegen des Signifikanztests
 - Festlegung des Signifikanzniveaus
 - Bestimmung der Nullhypothese
 - Vorgabe einer Testverteilung (in den nichtparametrischen Tests beim Testen einer Stichprobe)
 - Gewichtung von Beobachtungen
 - Behandlung von Bindungen

Verarbeitung:

1. Parametrische Tests
 - Vergleich zweier Mittelwerte (T-Test)
 - bei unabhängigen Ereignissen
 - bei abhängigen Ereignissen
 - Verhältnis der Varianzen
 - gleich
 - ungleich

- Vergleich von n-Mittelwerten
 - bei unabhängigen Ereignissen
 - bei abhängigen Ereignissen
 - Statistische Kennzahlen
 - Mittelwerte
 - Standardabweichung
 - Standardfehler des Mittelwertes
 - T-Statistik
 - t-Wert
 - Freiheitsgrade
 - Signifikanzniveau
 - F-Statistik
 - F-Wert
 - Freiheitsgrad jeder Untergruppe
 - Signifikanzniveau
- 2. Nichtparametrische Tests
 - Testen einer Stichprobe
 - Anpassungstests
 - Iterationstests
 - Test mehrerer unabhängiger Stichproben
 - Prä- und Posttests
 - Vorzeichentests
 - Rangskalentests
 - Tests auf Verteilungsparameter
 - Test mehrerer abhängiger Stichproben

Ausgabe:

1. Testergebnisse
2. Teststatistik

6.1.3 KORRELATIONEN

Eingabe:

1. Daten
 - Rohdaten
 - aufbereitete Daten
 - Matrizen
 - Skalierungsniveau der Daten
 - intervallskaliert
 - ordinal
2. Variablenfestlegung
 - maximale Anzahl an Variablennamen
 - maximale Anzahl an unabhängigen Variablen
3. Bestimmung der Korrelationsbeziehungen
 - maximale Anzahl an Schlüsselwörtern
 - maximale Anzahl an Trennzeichen
 - maximale Anzahl an Ordnungswerten (bei partieller Korrelation)
4. Methodenaufruf
 - Behandlung von fehlenden Werten
 - Festlegung der Signifikanztests
 - einseitig
 - zweiseitig
 - Festlegung des Signifikanzniveaus
 - Gewichtung von Beobachtungen
 - Bestimmung der Freiheitsgrade (bei kanonischer Korrelation)

Verarbeitung:

1. Pearson's Produkt-Moment-Korrelations-Koeffizient
 - Korrelationskoeffizient
 - Mittelwert
 - Standardabweichung
 - Summe
 - Irrtumswahrscheinlichkeit
 - einseitig
 - zweiseitig
2. Partieller Korrelationskoeffizient
 - Korrelationskoeffizient
 - Mittelwert
 - Standardabweichung
 - Freiheitsgrade
 - Irrtumswahrscheinlichkeit
 - einseitig
 - zweiseitig
3. Kanonische Korrelation
 - Kanonische Variable
 - Mittelwerte für unabhängige Variablen
 - Mittelwerte für abhängige Variablen
 - Korrelation zwischen den Variablen jeder Gruppe (unabhängig und abhängig) und der jeweiligen kanonischen Variablen
 - Eigenwerte
 - Fehlerquadratsumme
 - Signifikanztest
 - Lambda
 - Chi-Quadrat-Statistik
4. Spearman's Korrelationskoeffizient
 - Korrelationskoeffizient
 - Mittelwert
 - Standardabweichung
 - Median
 - Minimum
 - Maximum
 - Range
 - Irrtumswahrscheinlichkeit
 - einseitig
 - zweiseitig
 - maximale Anzahl der herangezogenen Beobachtungen
5. Kendall's Korrelationskoeffizient
 - Korrelationskoeffizient
 - Mittelwert
 - Standardabweichung
 - Median
 - Minimum
 - Maximum
 - Range
 - Irrtumswahrscheinlichkeit
 - einseitig
 - zweiseitig
 - maximale Anzahl der herangezogenen Beobachtungen

Ausgabe:

1. Tabelle der Korrelationskoeffizienten
 - seriell
 - in Matrizenform
 - auf neues Datenfile
2. Statistische Kennzahlen

6.1.4 REGRESSION

Eingabe:

1. Daten
 - Rohdaten
 - aufbereitete Daten
2. Variablenfestlegung
 - Bestimmung der unabhängigen Variablen
 - Bestimmung der abhängigen Variablen
 - maximale Anzahl an unabhängigen Variablen
 - maximale Anzahl an Variablen
3. Methodenaufruf
 - Angabe der "lags" (bei Autoregression)
 - maximale Anzahl an "lags"
 - Angabe eines Abbruchkriteriums (bei nicht-linearer Regression)
 - Festlegen eines Konvergenzkriteriums
 - Vorgabe der Iterationsschritte
 - maximale Anzahl an Schritten
 - Definitionsbereich für Parameter (bei nicht-linearer Regression)
 - Startwerte für das Lösungsverfahren
 - Angabe fixer Werte
 - Angabe eines Intervalls
 - Intervallgrenzen im Lösungsverfahren
 - Determinierung der einbezogenen Variablen (nicht bei Einfachregression)
 - Mindestanzahl an Variablen, die in der Regressionsberechnung berücksichtigt werden sollen
 - maximale Anzahl an Variablen, die in der Regressionsberechnung berücksichtigt werden sollen
 - Hierarchiestufen der einbezogenen Variablen
 - maximale Anzahl an Prioritäten
 - Bestimmung von Variablen, die in jede Berechnung miteinbezogen werden sollen
 - Signifikanzniveau (bei Mehrfachregression)
 - Kriterium für die Aufnahme einer Variablen in das Modell
 - Kriterium für das Verbleiben einer Variablen im Modell
 - Behandlung von fehlenden Werten
 - Berücksichtigung im Modell
 - Abbruch des Verfahrens
 - Ausschluß aus der Berechnung
 - Gewichtung der Parameter

Verarbeitung:

1. Einfachregression
2. Multiple Regression
 - Forward Selection
 - Backward Selection
 - Stepwise Method
 - Methode der maximalen R - Verbesserung
 - Methode der minimalen R - Verbesserung
 - simultane Berücksichtigung der Variablen
3. Nichtlineare Regression
 - Gradientenverfahren
 - Gauß-Newton-Methode
 - Marquardt-Methode
 - Gitterpunktsuche
4. Autoregression
 - Autokorrelationskoeffizienten
 - Autokovarianzen
5. Regression in einem System linearer Gleichungen
 - Methode der kleinsten Quadrate
 - einstufig
 - zweistufig
 - dreistufig
 - Maximum-Likelihood-Methode
 - Methode von Zellner ("seemingly unrelated regression")
6. Sonstige Regressionsverfahren
 - Regression mit Dummy-Variablen
 - gewichtete Regression
 - polynome Regression
 - Berechnung der Regression für jede Kombination zwischen abhängigen und unabhängigen Variablen
7. Statistische Kennzahlen
 - Korrelationskoeffizient und Bestimmtheitsmaß
 - Fehlerquadratsumme
 - F-Statistik zur Varianzanalyse
 - Test der Regressionskoeffizienten
 - Voraussagewerte

Ausgabe:

1. Regressions- und Statistiktabelle
2. Varianzanalyse

6.1.5 VARIANZANALYSE

Eingabe:

1. Daten
 - ungleiche Zellenbesetzungen erlaubt
 - leere Zellen erlaubt
2. Methodenaufruf
3. Variablenfestlegung
 - Anzahl der Faktoren
 - maximale Anzahl der Faktoren
 - abhängige Variable
 - maximale Anzahl der abh. Variablen
 - Kovariate
 - maximale Anzahl der Kovariate

- Verschachtelungen
- 4. Testaufruf
 - Zusammensetzung der Gruppen für t-Test
 - Parameter für a-posteriori-Test

Verarbeitung:

1. Einfaktorielle Varianzanalyse
2. Mehrfaktorielle Varianzanalyse
 - simultane Aufnahme
 - hierarchische Aufnahme
3. Kovarianzanalyse
4. Berechnung der Effekte
 - Haupteffekte
 - Wechselwirkungen
 - Kovariateffekte
5. Regressionskoeffizienten
6. Tests
 - Trendtests
 - Tests auf a-priori-Mittelwertunterschiede
 - Tests auf a-posteriori-Einzelmittelwertunterschiede
 - Varianzhomogenitätstests
7. Gruppenstatistiken

Ausgabe:

1. Varianzanalysetafel
2. Testergebnisse
3. Statistiken
4. Matrizen

6.1.6 FAKTORENANALYSE

Eingabe:

1. Daten
 - maximale Anzahl der Objekte
 - maximale Anzahl der Variablen
 - Dateien
 - Matrizen
 - Korrelationsmatrix zwischen Objekten
 - Korrelationsmatrix zwischen Variablen
 - Matrix der Faktorladungen
2. Methodenaufruf
3. Faktorenfestlegung
 - Anzahl der Faktoren
 - Bestimmung durch Festlegung eines Erklärungsanteils an der Eigenwert-Summe
 - minimaler Eigenwert der Korrelationsmatrix
4. Kommunalitäten
 - Kommunalitätsschätzungen
 - Abbruchkriterien für Endkommunalitätenschätzung
 - maximale Anzahl der Iterationsschritte
 - minimale Differenz zwischen zwei aufeinanderfolgenden Schätzungen

Verarbeitung:

1. Datenreduktion
 - Hauptkomponentenmethode
 - Hauptachsenfaktorenanalyse mit Kommunalitäteniteration
 - Rao's kanonische Faktorenanalyse
 - Alpha-Faktorenanalyse
 - Image-Faktorenanalyse
 - Maximum-Likelihood-Methode
2. Rotation
 - Varimax
 - Quartimax
 - Equimax
 - Oblique
 - Prärotation
 - Varimax
 - Quartimax
 - Equimax
3. Berechnung der Korrelationsmatrix
 - Eigenvektoren der Korrelationsmatrix
4. Varianzzerlegung
 - Kommunalitäten, Eigenwerte und Anteile der totalen und gemeinsamen Varianz
5. Faktorenmatrizen
 - unrotierte Faktorenmatrix
 - rotierte Faktorenmatrix
 - Transformationsmatrix
 - Faktorenwertkoeffizientenmatrix
 - Interfaktorenkorrelationen

Ausgabe:

1. Korrelationsmatrix
 - auch Inverse
 - auch Determinante
 - imageskalierte Korrelationsmatrix
 - Anti-Image-Korrelationsmatrix
 - Eigenwerte des Image
2. Faktorenwerte
3. Faktorenmuster
4. Matrizen

6.1.7 DISKRIMINANZANALYSE**Eingabe:**

1. Daten
 - Dateien
 - Matrizen
 - Stichprobenanteil für Analysephase
2. Methodenaufruf
 - Aufnahmekriterien
 - Anzahl der Iterationen
 - Anzahl der kanonischen Diskriminanzfunktionen bzw. kumulierter Prozentsatz der Eigenwerte bzw. Signifikanzniveau
 - Zugehörigkeitswahrscheinlichkeiten

3. Variablenfestlegung

- diskriminierende Variablen

Verarbeitung:

1. Analysephase

- simultaner Einschluß der diskrim. Variablen
- schrittweiser Einschluß
 - endgültige Aufnahme
 - Ausscheiden bei Verschlechterung der Gruppentrennung

2. Berechnungen

- Kovarianzen
 - Intra-Gruppen-Kovarianzmatrix
 - Gesamt-Kovarianzmatrix
 - Homogenitätstest der Intra-Gruppen-Kovarianzmatrizen
- Korrelationen
 - Intra-Gruppen-Korrelationsmatrix
 - Partielle Korrelationsmatrix
- Diskriminanzfunktionen
 - Varimax-Rotation der Diskr.funktionen
 - Aufnahmekennwerte
- Klassifikationsfunktionen
 - Koeffizienten
 - Varianzanalyse auf Gleichheit der Gruppenmittelwerte für jede diskriminierende Variable

3. Klassifikationsphase

- nearest neighbour
- k-nearest neighbour
- Test mit restlichen Fällen der Stichprobe

Ausgabe:

1. Diskriminanzwerte
2. Klassifikationsergebnisse
 - Trefferwahrscheinlichkeiten
 - Zugehörigkeitswahrscheinlichkeiten zu allen Gruppen
 - Fehlklassifikationen
 - Gruppendistanzen
3. Territoriumskarte
 - Gruppenflächen
 - Bildausgabe der Fälle
4. Matrizen

6.1.8 CLUSTERANALYSE

Eingabe:

1. Daten
 - Dateien
 - auch Binärdaten
 - maximale Anzahl der Variablen
 - maximale Anzahl der Objekte
 - Matrizen

2. Methodenaufruf

- Wahl des Distanz- bzw. Ähnlichkeitsmaßes
- Anzahl der Cluster
- Anzahl der zu berechnenden Nachbarn
- Clusterzuteilung

Verarbeitung:

1. Vorgangsweise

- nicht hierarchisch
 - seed-points
 - Anfangsverteilung
 - deterministisch
 - stochastisch
- hierarchisch
 - Linkage-Methoden
 - Centroid-Methoden
 - Fehlerquadratsummen- bzw. Varianzmethoden

2. Berechnungen

- Standardisierung
- Distanzmaße
 - Nachbarn
- Statistiken der Variablen
 - Eigenwerte, Eigenvektoren
- Statistiken der Cluster
- Tests
 - F-Werte für die Clustervariablen
 - t-Werte für die Clustervariablen

Ausgabe:

1. Protokoll

- Vorgangsprotokoll
- Statistiken
- Cluster-Mitglieder

2. Dendrogramm

- ausgedruckt
- unvollständiger Raster

6.2 DATEN- UND DATEI-HANDLING

Eingabe:

1. Variable

- maximale Anzahl pro Datei
- maximale Anzahl der Positionen
- Definitionsbereich numerischer Variabler
- Definitionsbereich binärer Variabler

2. Etiketten

- Etiketle der Variablen
 - maximale Anzahl der Positionen
- Etiketten für Ausprägungen
 - maximale Anzahl der Positionen

3. Formate

- fixe Eingabeformate
- freie Eingabeformate

4. Fehlende Werte

- maximale Anzahl der Kodierungen
- Unterscheidung zwischen Blank und Null

5. Dateinamen
 - maximale Anzahl der Positionen
 - maximale Anzahl der Unterdateien
 - maximale Anzahl der Positionen für Unterdateinamen
6. Zusatztext
 - maximale Anzahl der Positionen
7. Eingriff in die Speicherverwaltung

Aufbereitung von Dateien:

1. Sortieren
 - maximale Anzahl der Kriterien
2. Einbeziehungen
 - von Variablen
 - von Fällen
3. Änderungen
 - Ersetzen von Werten
 - Änderung von Variablennamen
 - Umordnung von Variablen
4. Berechnungen
 - Berechnung von neuen Variablen
 - Vorsorge für fehlende Werte
 - Bedingungen
 - Fallzählung
 - Lag-Verschiebung für Variable
 - Gewichtungsfaktoren für Variable
5. Hinzufügungen
 - von neuen Fällen
 - von neuen Variablen
 - von Unterdateien
6. Löschen
 - von Fällen
 - Variablen
 - Unterdateien
7. Zusammenfassungen
 - Erstellung von großen Archivdateien

Arbeiten mit Dateien:

1. Selektion
 - mittels logischem Ausdruck
 - mittels Zufallsstichprobe
2. Bedingungen und Schleifen
 - bedingte Verarbeitung
 - Schleifen
 - mit Bedingungen
3. Heranziehungsmöglichkeiten
 - Verwendung von Binärdateien
 - Definition von externen Ausgabedateien
 - Verwendung von Matrizen als Input
4. Multiplikationen
 - der Werte zweier Dateien
5. Standardisierung
6. Datenbankanschluß
 - Umwandlung von Dateien aus anderen Programmpaketen
 - Kopieren der Namen der Variablen
 - Kopieren der Etiketten der Variablen
 - Umwandlung alphabetischer Variabler
 - Kopieren der Etiketten der Werte
 - Formate

- 7. Kopieren
 - von Platte auf Band
 - von Platte auf Platte
 - von Band auf Platte
 - von Band auf Band
- 8. Dokumentation
 - der Datei
 - des Benutzers

6.3 AUSGABEGESTALTUNG

- 1. Ausgabeeinheiten
 - Drucker
 - Band/Magnetplatte
 - Datenfile
 - Terminal
 - Dialog-Verarbeitung
 - Batch-Verarbeitung
 - Ausgabeunterdrückung
 - selektiv
 - Weiterverwendung als Input für andere Prozedur
- 2. Charts
 - Histogramme
 - Gestaltungsform
 - vertikal
 - horizontal
 - Mehrfachbalken
 - Überlagerungsbalken
 - Variablenart
 - diskrete Variable
 - stetige Variable
 - Wertausgabe
 - absolut
 - prozentuell
 - kumuliert
 - mit Verteilung der Subgruppen
 - Flächendiagramme
 - Tortendiagramme
 - Flächenausschnitte
 - Umfang
 - Sterndiagramme
 - Flächenausschnitte
 - Umfang
 - Banddiagramme
 - Flächenausschnitte
 - Länge
 - Polaritätsprofile
 - Kriterienanzahl
 - Bewertungsskala
 - Koordinatensysteme
 - Art des Achsensystems
 - rechtwinklig
 - schiefwinklig
 - zweidimensional
 - dreidimensional

- Kurvenzeichnung
 - Streubilder
 - Linienführung
- Achsenskalierung
 - linear
 - logarithmisch
 - mit Interpolationen
 - mit Unterteilungen
 - Diagonale
 - Netzgitter
- Transformationsfunktionen
 - Drehung
 - Boxing
 - Verkleinerung
 - Vergrößerung
 - Zentrierung
- Legende
 - Zeichenerklärung
 - automatisch
 - nach Aufruf
 - selektive Unterdrückbarkeit
 - Prozedurhinweise
 - Methodenerklärung
 - Ergebnishinweise
 - Warnung bei Verletzung der Grundannahmen
 - Interpretation
- 3. Report Writing
 - Textgestaltung
 - Textblöcke
 - links ausgerichtet
 - rechts ausgerichtet
 - Schriftarten
 - Buchstabentypen
 - Strichart
 - Seitengestaltung
 - Seitenkopf
 - Format
 - fix
 - frei festlegbar
 - Zeilenabstand
 - Zeilen pro Seite
 - Spalten pro Seite
 - Numerierung
 - Jobdokumentation
 - Hinweise auf den Job
 - Fehlermeldungen
 - Unterdrückung
 - Ergebnisausgabe
 - seriell
 - Tabelle
 - Matrize

Exkurs 1:

ÜBERPRÜFUNG VON FRAGEBOGENITEMS

Eingabe:

1. Daten
 - Dateien
 - Matrizen
 - Korrelationsmatrix
 - Kovarianzmatrix
2. Bestimmung der Skalen
 - maximale Anzahl an Skalen
3. Festlegung der Variablen, deren Zusammenhang überprüft werden soll
 - maximale Anzahl an Variablen
 - Festlegung der Anordnung der Variablen bei der Ausgabe
4. Festlegung der Trennstellen
5. Methodenaufruf
 - Behandlung von fehlenden Werten
 - Berücksichtigung in der Berechnung
 - fallweiser Ausschluß

Verarbeitung:

1. Dichotomisierung der Variablen
2. Guttman-Skala
3. Statistische Kennzahlen für Guttman-Skala
 - nach der Goodenough-Technik
 - Interkorrelationsmaß
 - Reproduzierbarkeit
 - minimale Rand-Reproduzierbarkeit
 - prozentuelle Verbesserung
 - Skalierbarkeitskoeffizient
 - nach dem Verfahren von Proctor
4. Reliabilitätskoeffizienten
 - Konsistenzkoeffizient
 - Split-Half-Reliabilität
 - Guttman's untere Grenze für die wahre Reliabilität
 - Maximum-Likelihood-Schätzungen der Testreliabilität
5. Statistische Maßzahlen der Items
 - Mittelwerte
 - Standardabweichungen
 - Kovarianzmatrix
 - Interkorrelationsmatrix
 - Varianz
 - Trennschärfekoeffizient
 - multiple Korrelationen
 - Kennzahlen von Subtests
6. Tests für das Verhältnis der Items zueinander
 - Test der Itemmittelwerte
 - Tukey's Test
 - Varianzanalyse

Ausgabe:

1. Skalogramm für Guttman-Analyse
 - Randverteilungen
 - Häufigkeiten der richtigen und falschen Antworten pro Skalenwert und Variable
2. Reliabilitätstabelle
3. Varianzanalyse
4. Testergebnisse

QUELLENVERZEICHNIS

- Berekoven / Eckert / Ellenrieder, Marktforschung -
Methodische Grundlagen und praktische Anwendung,
Wiesbaden 1977
- Francis, I., A Taxonomy Of Statistical Software, in:
Barritt, M.M. / Wishart, D.W., Comstat 1980 -
Proceedings in Computational Statistics, S.43-54,
Wien 1980
- Frank, J., Standard-Software, Köln 1977
- Hansen, H.R. (1981), Stand und Entwicklungstendenzen
auf dem Markt für Standardanwendungssoftware, in:
Output 2/1981 (erscheint demnächst)
- Hansen, H.R. (1979), Möglichkeiten und Probleme beim
Einsatz von Standardanwendungssoftware, in:
IBM Nachrichten, Mai 1979, S.149-155
- Hansen, H.R. / Bauer, H., Fallstudie Computeraus-
wahl mit Hilfe eines Dialogprogrammsystems zur
Nutzwertanalyse (DIANA), Arbeitsbericht zum
Tätigkeitsbereich Wirtschaftsinformatik,
Wien 1979
- Hansen, H.R. / Maier B., Grundzüge der Datenverarbei-
tung und Anwendungsprogrammentwicklung, München 1980
- Infratest / MBP, Studie über Stand und Entwicklung des
Einsatzes von mehrfach verwendbarer Anwendungssoft-
ware in der gewerblichen Wirtschaft, Bonn -
München 1976
- Jordan, C., Datentabellen in der Systemplanung, Stutt-
gart 1976
- Kanngiesser, J., Anwendungssoftware - DV-Sorgenkind der
80er Jahre, in: Online 12/1980, S.1006-1010
- Komarnicki, J., Verfehlte DV-Marktstruktur als Ursache
der Software-Krise, in: Online 1/1978, S.42-46
- Kotler, P., Marketing Management - Analysis, Planning
and Control, 3rd edition, London 1976
- Nie, E.H. / Hull, C.H. / Jenkins, J.K. / Steinbrenner, K./
Bendt, D.H., SPSS, Statistical Package For The Social
Sciences, 2nd edition, New York 1975

Nomina Gesellschaft für Wirtschafts- und Verwaltungs-
register (Hrsg.), ISIS Software Report, München
(sämtliche Ausgaben seit 1979)

Steinhausen, D. / Langer, K., Clusteranalyse, Einführung
in Methoden und Verfahren der automatischen Klassifi-
kation, Berlin - New York 1977

Wedekind, H., Systemanalyse, 2. Auflage, München -
Wien 1976

Zimmermann, G., Qualitätsmerkmale von Standardsoftware und
Möglichkeiten ihrer Beurteilung, in: Online 4/1978,
S.300-303, Online 5/1978, S.419-422, Online 6/1978,
S.504-507

o.V., Die neue Superkraft, in: Computer Magazin 8/1980,
S.37-51