

ZEITSCHRIFT  
FÜR SOZIAL  
PSYCHOLOGIE

HERAUSGEBER  
HUBERT FEGER  
C.F. GRAUMANN  
KLAUS HOLZKAMP  
MARTIN IRLE

BAND 11 1980 HEFT 3

VERLAG HANS HUBER  
BERN STUTTGART WIEN

# Zeitschrift für Sozialpsychologie

1980, Band 11, Heft 3

---

## INHALT

---

<i>Zu diesem Heft</i> .....	139
<b>Theorie und Methoden</b>	
LANGEHEINE, R.: Multivariate Hypothesentestung bei qualitativen Daten	140
<b>Empirie</b>	
BRANDSTÄTTER, H., STOCKER-KREICHGAUER, G. & FIRCHAU, V.: Wirkung von Freundlichkeit und Argumentgüte auf Leser eines Diskussionsprotokolls. Ein Prozeßmodell .....	152
MÜLLER, G.F.: Interpersonales Konfliktverhalten: Vergleich und experimentelle Untersuchung zweier Erklärungsmodelle .....	168
BIERHOFF, H.W.: Naive psychologische Theorien und Eigenschaften als Funktion des Interaktionsmusters der Stimulusperson .....	181
<b>Diskussion</b>	
LEUTNER, D.: Einige Anmerkungen zu Guttmans «First Law of Attitude» .....	189
<b>Literatur</b>	
Neuerscheinungen .....	195
Titel und Abstracta .....	198
<b>Nachrichten und Mitteilungen</b> .....	201
<b>Autoren</b> .....	205

---

Copyright 1980

Verlag Hans Huber Bern Stuttgart Wien

Herstellung: Satzatelier Paul Stegmann, Bern

Printed in Switzerland

Library of Congress Catalog Card Number 78-126626

Die Zeitschrift für Sozialpsychologie wird in *Social Sciences Citation Index (SSCI)* und *Current Contents/ Social and Behavioral Sciences* erfaßt.

## Zu diesem Heft

Im Teil «Theorie und Methoden» findet sich von Herrn LANGEHEINE ein Beitrag, der sich mit der Analyse multipler Zusammenhänge bei nominalskalierten Daten befaßt. Die Methodenentwicklung in diesem Bereich wurde insbesondere in den Sozialwissenschaften i. e. S. vorangetrieben und hat die Psychologie noch nicht voll erreicht. Allerdings ist eine Variante durch die Arbeiten von LIENERT und KRAUTH bekannt geworden: Die Konfigurations-Frequenz-Analyse. Auf diese Analyse-methode geht Herr LANGEHEINE kritisch ein; zu der von ihm geäußerten Kritik wird im nächsten Heft dieser Zeitschrift Herr KRAUTH Stellung nehmen.

Von Herrn BRANDSTÄTTER und seinen Mitarbeitern aus dem ehemaligen Augsburger Projekt zur Medienforschung liegt ein Beitrag vor, der durch die zur Zeit von der DFG unternommenen Bemühungen zur Förderung der Medienforschung besondere Aktualität erhält.

Die Arbeit von Herrn MÜLLER über Interpersonelles Konfliktverhalten geht in Richtung eines theorietestenden Experimentes, eine Entwicklung, die bei dem gegenwärtigen Stand der experimentellen Erforschung sozialer Konflikte sehr wünschenswert ist.

Die Untersuchung von Herrn BIERHOFF führt die Untersuchungen des Autors zur Attributionstheorie fort.

Schließlich greifen wir mit dem Diskussionsbeitrag von Herrn LEUTNER jene Richtung der Sozialpsychologie auf, die schon seit längerem im Zusammenhang mit der mathematischen Meßtheorie betrieben wird und die in dem Bereich Einstellungsmessung Kenner und Liebhaber gefunden hat.

Alle Leser sind herzlich eingeladen, sozialpsychologisch relevante Informationen für den Nachrichtenteil dieser Zeitschrift der Redaktion mitzuteilen.

H. Feger



# Theorie und Methoden

## Multivariate Hypothesentestung bei qualitativen Daten\*

ROLF LANGEHEINE

Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel

Zumindest im deutschen Sprachraum ist die Konfigurationsfrequenzanalyse (KFA, KRAUTH & LIENERT, 1973) zu einem populären Ansatz für die Analyse qualitativer Daten geworden. In dieser Arbeit wird auf einige Schwächen der KFA eingegangen: (1) Da die Zerlegung des Gesamt Chi-Quadrats im symmetrischen Ansatz lediglich aufgrund einer Reihe von Modelltests nach der Hypothese der totalen Unabhängigkeit erfolgt, besteht die Möglichkeit, daß bedeutsame strukturelle Eigenschaften der Daten nicht adäquat erfaßt werden. (2) Die KFA ist für die asymmetrische Fragestellung unbrauchbar wegen Nichtadditivität der Chi-Quadrat Komponenten. (3) Wird die Angemessenheit der KFA zur Identifikation von Typen infragegestellt, auf die aus einem mit den Daten nicht kongruenten Modell geschlossen wird. Log-lineare Modelle sind dagegen nicht nur frei von den Schwächen (1) und (2), sondern lassen Schlußfolgerungen über die Struktur der Daten nur aus einem mit den Daten verträglichen Modell zu. Die Konsequenzen werden anhand von Reanalysen publizierter Daten demonstriert.

Configuration frequency analysis (CFA, KRAUTH & LIENERT, 1973) has become a popular approach to the analysis of qualitative data at least in German speaking areas. This paper is devoted to some limitations of CFA: (1) Since the partition of the total chi-square in the symmetric approach is exclusively due to a series of tests according to the hypothesis of total independence, possibly substantive structural characteristics may not be represented adequately. (2) CFA fails to be a useful asymmetric approach due to nonadditivity of the chi-square components in this case. (3) The adequacy of CFA in identifying types is questioned, since types are inferred from a model not congruent with the data. Log-linear models, on the other hand, are not only free from limitations (1) and (2) but allow inferences as to the structure of the data only from a model fitting the data. Consequences are demonstrated by reanalyzing some previously published data.

### 1. Einführung

Die KFA (Konfigurations-Frequenz-Analyse, KRAUTH & LIENERT, 1973) ist offensichtlich zumindest im deutschen Sprachraum zu einem populären Ansatz geworden, multiple Zusammenhänge zwischen nominal skalierten Variablen zu analysieren. KRÜGER (1979) hat kürzlich empfohlen, der KFA auch bei Ordinaldaten den Vorzug gegenüber der von SCHULZE (1978) vorgestellten Hierarchischen Rangvarianzanalyse (HRVA) zu geben. Er begründet dies mit zwei Argumenten: 1) Die HRVA arbeitet nicht echt-, sondern nur quasi-multivariat. 2) Die Reihenfolge, in der die

Variablen spezifiziert werden, hat Einfluß auf die Ergebnisse. Beide Einwände treffen zu, wenngleich der zweite Fall kein Spezifikum der HRVA sondern typisch ist für schrittweise Prozeduren.

In dieser Arbeit soll nun gezeigt werden, daß die offensichtlich bekanntesten und beliebtesten Varianten der KFA, die ASA (Assoziations-Struktur-Analyse oder hierarchische KFA) und die PKFA (Prädiktions-KFA oder Interaktionsstrukturanalyse), lediglich zwei Spezialfälle eines allgemeineren Modells sind, nämlich GOODMAN'S «general log-linear model» (GOODMAN, 1970, 1971, 1972a, 1972b). Wie bei der ASA bzw. der PKFA ist auch in GOODMAN'S Modell eine symmetrische Betrachtungsweise (assoziationsanalytischer Ansatz) und eine asymmetrische Betrachtungsweise (regressions- bzw. varianzanalytischer Ansatz) möglich. In Abschnitt 3 wird jeweils

\* Für Kommentare und Vorschläge zu einer ersten Version dieser Arbeit bin ich GERHARD ARMINGER und PETER KAPPELHOFF dankbar.

das allgemeine Modell für die symmetrische und asymmetrische Fragestellung vorgestellt und anhand von Reanalysen publizierter Ergebnisse gezeigt, daß die aufgrund der KFA gezogenen Schlußfolgerungen – Interpretation von Typen – in der Regel fragwürdig sind. Schließlich wird für ein Beispiel mit zwei unabhängigen und zwei abhängigen Variablen demonstriert, wie GOODMANS Modell ebenfalls ein Analogon zur multivariaten Varianzanalyse bei qualitativen Daten bietet.

Zuvor sei darauf hingewiesen, daß es zur Lösung des Problems – Bestimmung multipler Zusammenhänge bei nominal skalierten Daten – verschiedene Ansätze gibt, von denen hier wiederum nur einige angesprochen werden. Historische Überblicke über die ersten Versuche zur Lösung des Problems finden sich bei LANCASTER (1960, 1969a), der selbst mit seiner allgemeinen Chi-Quadrat-Zerlegung eine Reihe wichtiger Beiträge geleistet hat (z. B. LANCASTER, 1951, 1960, 1969a, 1969b). SUTCLIFFE (1957) hat LANCASTERS Ansatz in die psychologische Literatur eingeführt. Die Darstellung bei WINER (1970), wiederum, beruht auf der von SUTCLIFFE. Die Varianten der KFA (KRAUTH & LIENERT, 1973), schließlich, sind identisch mit dem Ansatz von LANCASTER mit dem Unterschied, daß sich KRAUTH & LIENERT darüber hinaus für die Identifikation von überzufällig oder unterzufällig auftretenden Typen interessieren.

Ein zweiter Strang findet sich in dem informationstheoretischen Ansatz von KULLBACK und Mitarbeitern (KU & KULLBACK, 1968; KU et al., 1971; KULLBACK, 1974). Obwohl dieser Ansatz in der Regel zu gleichen Ergebnissen führt wie eine Analyse nach GOODMANS log-linearem Modell, bestehen doch gewisse Unterschiede (vgl. z. B. BISHOP et al., 1975; GOODMAN, 1970, 1971).

Eine weitere Möglichkeit bietet der nach GRIZZLE, STARMER & KOCH (1969) benannte GSK-Ansatz gewichteter Regression. Wie der Ausdruck «Regression» bereits besagt, lassen sich über den GSK-Ansatz lediglich asymmetrische Fragestellungen behandeln, dies jedoch in sehr flexibler Weise, wobei die Testung hierarchischer wie nicht-hierarchischer Modelle (vgl. Abschnitt 3.1) möglich ist.

Gegenüber dem GSK-Ansatz ist in GOODMANS Modell sowohl die symmetrische als auch die asymmetrische Betrachtungsweise möglich. Neben den damit angedeuteten Analysestrategien

lassen sich daher in Analogie zur Pfadanalyse bei metrischen Daten simultane Gleichungssysteme testen, wie GOODMAN (1972a, 1973a, 1973b, 1979; vgl. auch KÜCHLER, 1978, 1979) gezeigt hat. Schließlich sei erwähnt, daß GOODMAN (1974a, 1974b, 1978) seinen Ansatz zur Analyse latenter Strukturen erweitert hat, wobei einige Variablen als manifest und andere als latent angesehen werden.

Obwohl insbesondere auf die beiden letzteren Ansätze verschiedentlich in der psychologischen Literatur aufmerksam gemacht wurde (MEREDITH et al., 1974; SHAFFER, 1973; SMITH, 1976), zeigt eine Durchsicht psychologischer Zeitschriften – sowohl deutscher wie englischsprachiger –, daß sie offensichtlich wenig bekannt sind. Es sei daher auf folgende Texte verwiesen: BISHOP et al. (1975), EVERITT (1977), FIENBERG (1978), KÜCHLER (1979), LANGEHEINE (1980) und REYNOLDS (1977a, 1977b – hier findet der Leser eine kurze Einführung in log-lineare Modelle).

## 2. Methoden oder Modelle?

Da der Begriff «Modell» in dieser Arbeit eine entscheidende Rolle spielt, soll hierauf zunächst kurz eingegangen werden. Die Ausbildung von Psychologen beginnt in der Regel u. a. mit einer Veranstaltung mit etwa folgendem Titel: «Quantitative Methoden der Psychologie». Bleibt der Veranstaltungsleiter dem Titel treu, so wird der Student wenig Modelldenken lernen, obwohl er bereits mit dem arithmetischen Mittel und dem Median zwei Modelle kennenlernt, mit denen die zentrale Tendenz einer Verteilung charakterisiert werden kann. Die Wege (Methoden, Verfahren) zur Berechnung jeder dieser Maßzahlen können sehr unterschiedlich sein, das Ergebnis ist jedoch jeweils methodenunabhängig. Beide Maßzahlen werden dagegen in der Regel nicht identisch sein. Die Ergebnisse sind damit modellabhängig. Es sei auf die ausführlichere Diskussion bei KRIZ (1973) verwiesen.

Im Fall der Kontingenztabellenanalyse ist ein Modell eine Hypothese über Verteilungen von und Beziehungen zwischen Variablen. Anders ausgedrückt: Mit einem Modell werden Annahmen (Hypothesen) darüber spezifiziert, wie die Zellhäufigkeiten zustandekommen könnten. Die Begriffe «Modell» und «Hypothese» werden da-

her synonym verwendet (REYNOLDS, 1977b). In der Tat gibt es bereits für eine  $2 \times 2$  Kontingenztabelle vier mögliche Modelle (vgl. z. B. REYNOLDS, 1977b). Ziel einer log-linearen Analyse ist daher die Identifikation eines Modells, das die Variation in den Zelhäufigkeiten erklärt, und zwar möglichst einfach (Ökonomieprinzip). Ob ein Modell dies leistet (mit den Daten kongruent ist, sie fittet), läßt sich über einen Modelltest bestimmen.

Obwohl ein Modell aufgrund einer inhaltlichen Theorie vor der Datenerhebung spezifiziert werden sollte, zeigt die Praxis in der Regel auch hier, daß durch ex post Testung verschiedener Modelle ein zufriedenstellendes Modell gesucht wird. Der Modellfindungsprozeß läuft dabei in mehreren Schritten ab (vgl. REYNOLDS, 1977b):

- 1) Spezifiziere ein Modell, das die beobachteten Häufigkeiten erklären könnte.
- 2) Berechne erwartete Häufigkeiten unter der Annahme, daß das Modell wahr ist.
- 3) Entscheide durch Vergleich von beobachteten und erwarteten Häufigkeiten, ob das Modell akzeptabel ist.
- 4) Fällt der Modelltest nach Schritt 3 positiv aus, so gehe nach Schritt 5. Andernfalls gehe nach Schritt 1 und spezifiziere ein neues Modell.
- 5) Schätze die Modellparameter und übersetze sie in substantielle Aussagen.

Diese Schritte machen zugleich deutlich, daß die Ableitung substantieller Schlußfolgerungen nur aus einem Modell sinnvoll ist, das mit den Daten kongruent ist. Gilt ein Modell nicht, so führt die Interpretation von Parametern (Effekten) zu Fehldiagnosen bzw. Fehlprognosen. Ein Beispiel – in Anlehnung an KRIZ (1973): Der Leser ist mit seinem PKW auf einer Reise zu einer Tagung in einer ihm bisher unbekanntem Stadt. Er hat einen Stadtplan (Modell der Stadt) bei sich, auf dem die wesentlichen Straßen falsch eingezeichnet sind (das Modell gilt – wenigstens zum Teil – nicht). Was passiert?

### 3. Adäquate Modelle für drei Fragestellungen

FEGER (1978, p. 306) geht es in Kapitel VII seines Buches um «... die Frage, ob sich Typen des Konfliktlebens nachweisen lassen». Analysiert wurden zwei Stichproben, für die in den am Median

dichotomisierten Variablen Konfliktstärke (A)<sup>1</sup>, Konfidenz (B) und Wichtigkeit (C) Daten für 8 bzw. 7 Konfliktsituationen vorlagen. Da keine der Variablen als abhängig von den restlichen angesehen wird, wurde die KFA-Analyse folgerichtig nach dem assoziationsanalytischen Ansatz durchgeführt. In Abschnitt 3.2 wird über Reanalysen für die 8 Situationen der ersten Stichprobe berichtet.

KRÜGER (1979, p. 94) entwickelt seine Überlegungen an einem Datensatz, bei dem «... die Abhängigkeit der Schulnoten von Schülermerkmalen an Volksschülern untersucht» wird. Die Deutschnote (N) wird als Prädikand angesehen, während dem Geschlecht (G), dem Sozialstatus des Vaters (S) sowie der Leistung im Rechtschreibtest (R) der Status von Prädiktoren zugewiesen wird. Es handelt sich somit um eine asymmetrische Fragestellung, die zunächst mit der entsprechenden PKFA angegangen wird. Zur Auswahl optimaler Prädiktoren empfiehlt KRÜGER dann jedoch die ASA, da die PKFA die Interaktion zwischen Prädiktoren nicht berücksichtigt. Antworten auf eine asymmetrische Fragestellung sollten jedoch mit einem entsprechenden Modell gegeben werden. In Abschnitt 3.4 wird daher gezeigt, welche Schlußfolgerungen sich aus einer Logitanalyse (Spezialfall des allgemeinen log-linearen Modells) ableiten lassen.

Der dritte hier reanalytierte Datensatz findet sich ebenfalls bei FEGER (1978, Tab. 7.11). Von den vier Variablen dieser Tabelle werden Wichtigkeit (A)<sup>2</sup> sowie die relative Stärke widerstreitender Tendenzen (D) als unabhängige Variablen und Konfliktstärke (B) sowie Entscheidungszeit (C) als abhängige Variablen bezeichnet (FEGER, 1978, p. 320). Dieser Konstellation entspricht das Design einer multivariaten 2-faktoriellen Varianzanalyse. Analysiert wird von FEGER jedoch nach dem Modell der ASA, um die Vorhersage von aus theoretischen Überlegungen abgeleiteten Typen zu prüfen. In Abschnitt 3.5 wird daher für diese Daten eine multivariate Varianzanalyse mit GOODMANS Modell durchgeführt.

<sup>1</sup> FEGERs Kurzbezeichnungen sind K, C und W.

<sup>2</sup> A, B, C und D entsprechen FEGERs Kurzbezeichnungen W, KS, EZ und RS.

3.1 *Symmetrische Fragestellung: Assoziations-analytischer Ansatz*

Im Fall einer 3-Weg Kontingenztabelle mit den Variablen A, B und C mit den Kategorien  $i = 1, 2, \dots, I, j = 1, 2, \dots, J$  und  $k = 1, 2, \dots, K$  suchen wir ein Modell für die Wahrscheinlichkeit  $p_{ijk}$ , mit der sich ein Merkmalsträger (Person usw.) aus der Population in der Zelle  $ijk$  befindet (wobei  $\sum_{ijk} p_{ijk} = 1$ ). Log-lineare Modelle arbeiten entweder mit diesen  $p_{ijk}$  (genauer: Schätzwerten aufgrund der Stichprobe) oder den unter einem Modell erwarteten Häufigkeiten  $m_{ijk}$  oder den natürlichen Logarithmen der erwarteten Häufigkeiten  $\mu_{ijk} = \ln m_{ijk}$ . Im Prinzip ist es gleichgültig, für welches Vorgehen man sich entscheidet, da die Modelle ineinander überführbar sind. Endziel ist es, die  $m_{ijk}$  oder  $\mu_{ijk}$  als Funktion einer Anzahl von Parametern auszudrücken (vgl. Schritt 5, Abschnitt 2).

Berücksichtigt man alle möglichen Parameter für eine 3-Weg Tabelle, so lautet das saturierte log-lineare Modell (additives Modell für die Logarithmen der erwarteten Häufigkeiten und somit multiplikatives Modell für die erwarteten Häufigkeiten):

$$(1) \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

In Analogie zur Varianzanalyse entsprechen die  $\lambda$  Parameter einer Konstanten ( $\lambda = \text{grand mean}$ ), drei Haupteffekten ( $\lambda_i^A, \lambda_j^B$  und  $\lambda_k^C$ ), drei 2er Interaktionseffekten ( $\lambda_{ij}^{AB}, \lambda_{ik}^{AC}$  und  $\lambda_{jk}^{BC}$ ) und

einem 3er Interaktionseffekt ( $\lambda_{ijk}^{ABC}$ ). Wie bei der Varianzanalyse muß, um eine eindeutige Lösung zu erhalten, reparametrisiert werden, z. B. mit

$$(2) \sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_k \lambda_k^C = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = \dots =$$

$$\sum_i \lambda_{ijk}^{ABC} = \sum_j \lambda_{ijk}^{ABC} = \sum_k \lambda_{ijk}^{ABC} = 0.$$

(1) wird als saturiertes Modell bezeichnet, da die Anzahl unabhängiger Parameter gleich der Anzahl der Zellen der Tabelle ist. Verschiedene unsaturierte Modelle lassen sich aus (1) dadurch spezifizieren, daß entsprechende Parameter des saturierten Modells gleich null gesetzt werden. In Tabelle 1 finden sich 11 solcher Modelle. Für Modell H1 wird z. B. angenommen, daß die 3er-Interaktion ohne Bedeutung ist. Modell H5 entspricht der Annahme der totalen Unabhängigkeit der 3 Variablen. Alle Interaktionsparameter sind daher auf Null gesetzt.

Alle in dieser Arbeit vorgestellten Modelle sind zugleich hierarchische Modelle<sup>3</sup>. Das Hierarchieprinzip läßt sich wie folgt beschreiben: Ist in einem Modell ein Parameter mit einem Satz S von Superskripten enthalten, so enthält das Modell automatisch alle Parameter mit Untermengen von S als Superskripten. Beispiel: Das Modell H3, in Tabelle 1 enthält u. a. den Parameter  $\lambda_{ij}^{AB}$ . Damit enthält es zugleich die Parameter  $\lambda_i^A$  und  $\lambda_j^B$ . Entspre-

<sup>3</sup> Hinweise zur Spezifikation nicht-hierarchischer log-linearer Modelle finden sich bei LANGEHEINE (1979).

Tab. 1: Mögliche Hypothesen zu einer 3-Weg Kontingenztabelle.

Hypothese	Modell	angepaßte Randverteilungen	Freiheitsgrade
H0	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$	(ABC)	0
H1	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$	(AB), (AC), (BC)	(I-1)(J-1)(K-1)
H2 <sub>1</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$	(AB), (AC)	I(J-1)(K-1)
H2 <sub>2</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC}$	(AB), (BC)	J(I-1)(K-1)
H2 <sub>3</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$	(AC), (BC)	K(I-1)(J-1)
H3 <sub>1</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$	(AB), (C)	(IJ-1)(K-1)
H3 <sub>2</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC}$	(AC), (B)	(IK-1)(J-1)
H3 <sub>3</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$	(BC), (A)	(JK-1)(I-1)
H4 <sub>1</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	(AB)	IJ(K-1)
H4 <sub>2</sub>	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_k^C + \lambda_{ik}^{AC}$	(AC)	IK(J-1)
H4 <sub>3</sub>	$\mu_{ijk} = \lambda + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$	(BC)	JK(I-1)
H5	$\mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C$	(A), (B), (C)	IJK-I-J-K+2

chend diesem Hierarchieprinzip werden log-lineare Modelle daher häufig nur durch die Minimalmenge der angepaßten Randverteilungen charakterisiert (vgl. Tabelle 1).

Einige Hinweise noch zu den Schritten des Modellfindungsprozesses (vgl. Abschnitt 2). Die unter einem Modell erwarteten Häufigkeiten lassen sich teilweise aus den Randverteilungen der Daten berechnen. Nicht in jedem Fall existieren jedoch explizit angebbare Formeln. Trifft dies zu (z. B. für die Hypothese H1 in einer 3-Weg Tabelle), so erfolgt die Berechnung über ein iteratives Verfahren (vgl. z. B. BISHOP et al., 1975; GOODMAN, 1970; REYNOLDS, 1977a).

Für den Modelltest (Schritt 3) bieten sich zwei Testgrößen an:

$$(3) P \text{ Chi}^2 = \sum_{ijk} (n_{ijk} - m_{ijk})^2 / m_{ijk} \text{ und}$$

$$(4) LR \text{ Chi}^2 = 2 \sum_{ijk} n_{ijk} \cdot \ln(n_{ijk} / m_{ijk}).$$

(3) ist das bekannte Pearson  $\text{Chi}^2$ , (4) ist die Likelihood Ratio Testgröße, die gegenüber (3) einige Vorteile bietet (vgl. z. B. BISHOP et al., 1975) und daher auch in dieser Arbeit ausschließlich verwendet wird. Die Anzahl der Freiheitsgrade (df) ergibt sich aus der Differenz Gesamtzahl der Zellen minus Anzahl der voneinander unabhängigen Parameter eines Modells. Explizite Formeln finden sich in Tabelle 1. Ein signifikantes  $\text{Chi}^2$  impliziert, daß das Modell nicht mit den Daten kongruent ist. Zur Findung eines befriedigenden Modells muß daher um zusätzliche Parameter erweitert werden.

Liegt kein explizit aus theoretischen Überlegungen abgeleitetes Modell vor und ist die Anzahl

der Variablen groß (4 und mehr), so kann der Suchprozeß nach einem befriedigenden Modell außerordentlich aufwendig sein. Es sei daher auf BENEDETTI & BROWN (1976) verwiesen, die eine vergleichende Untersuchung zur Effizienz von 13 Strategien durchgeführt haben.

Die Schätzung der Modellparameter (Schritt 5) erfolgt aus den Logarithmen der unter einem Modell erwarteten Häufigkeiten. Die  $m_{ijk}$  müssen daher größer Null sein. BISHOP et al. (1975) behandeln ausführlich, wie sich das Problem von Nullzellen in den Daten lösen läßt.

Schließlich gibt es zwei Möglichkeiten, Aussagen über die Bedeutsamkeit von (Haupt- und Interaktions-) Effekten in Modellen zu machen. Der erste Ansatz (vgl. z. B. GOODMAN, 1970, 1971) bedient sich der Möglichkeit, die Standardabweichung der  $\lambda$  zu schätzen und mit den standardisierten Werten von  $\lambda$  (d. h. mit  $\lambda/s_\lambda$ ) anhand der Normalverteilungstabelle die Hypothese zu testen daß ein bestimmtes  $\lambda$  gleich Null ist. Der zweite Ansatz beruht auf dem Konzept der Chi-Quadrat Zerlegung (vgl. u. a. BISHOP et al., 1975; GOODMAN, 1970; REYNOLDS, 1977a). Z. B. unterscheiden sich die Modelle H1 und H2, in Tabelle 1 lediglich in dem Parameter  $\lambda_{jk}^{BC}$ . Unter der Bedingung, daß H1 gilt, läßt sich der Effekt BC daher aus der Differenz der entsprechenden LR Chi-Quadrate quantifizieren, mit anderen Worten:

$$(5) \text{Chi}^2(\text{BC} | \text{H1}) = \text{Chi}^2(\text{H2}_1) - \text{Chi}^2(\text{H1}).$$

Tab. 2: Ergebnisse der Modelltests für 8 Situationen.

Hypo- these	df	Sit. A		Sit. B		Sit. C		Sit. D		Sit. E	
		LRChi <sup>2</sup>	p								
H1	1	.38	>.5	2.22	.137	.06	>.5	.01	>.5	.14	>.5
H2 <sub>1</sub>	2	9.58	.008	2.83	.243	7.39	.025	13.92	.001	.25	>.5
H2 <sub>2</sub>	2	.61	>.5	2.22	.330	1.28	>.5	.11	>.5	6.10	.047
H2 <sub>3</sub>	2	27.88	.000	38.42	.000	55.00	.000	23.13	.000	14.07	.001
H3 <sub>1</sub>	3	13.43	.004	2.99	.392	19.12	.000	17.74	.001	6.26	.099
H3 <sub>2</sub>	3	40.70	.000	39.13	.000	72.84	.000	40.75	.000	14.22	.003
H3 <sub>3</sub>	3	31.73	.000	38.58	.000	66.72	.000	26.94	.000	20.07	.000
H4 <sub>1</sub>	4	13.49	.009	3.04	>.5	19.43	.001	18.04	.001	6.26	.180
H4 <sub>2</sub>	4	40.86	.000	39.50	.000	72.85	.000	41.06	.000	14.28	.007
H4 <sub>3</sub>	4	32.48	.000	38.58	.000	66.78	.000	27.25	.000	20.12	.001
H5	4	44.55	.000	39.35	.000	84.57	.000	44.57	.000	20.22	.001