Kreuztabellenanalyse und Analyse von Individualdaten mit GLIM

Andreß, Hans-Jürgen

KREUZTABELLENANALYSE UND ANALYSE VON INDIVIDUALDATEN MIT GLIM

1. Einleitung und Problemstellung

Es ist das Verdienst von ARMINGER (1982, 1983), den GLIM-Ansatz [1)] von NELDER & WEDDERBURN (1972) einer breiteren wissenschaftlichen Öffentlichkeit zugänglich gemacht zu haben. Dieser Ansatz wurde bisher in der sozialwissenschaftlichen Diskussion wenig beachtet, obwohl er ein allgemeines Modell sozialwissenschaftlicher Datenanalyse formuliert, unter das sich die verschiedensten Methoden subsumieren lassen. Ein Grund für die geringe Verbreitung ist sicherlich die stark formalstatistische Formulierung dieses Ansatzes, die den Zugang für den Anwender statistischer Methoden erschwert und innerhalb derer die vielen Submodelle des GLIM-Ansatzes nicht unmittelbar evident sind.

ARMINGER beschränkt sich in seiner Darstellung (1983) auf die Modelle, die dem Anwender aus der multivariaten Kreuztabellenanalyse bekannt sind. Er kann dabei zeigen, daß

- verschiedene Ansätze der Kreuztabellenanalyse (log-lineare Modelle, GSK-Ansatz), die in der Vergangenheit zum Teil sehr kontrovers diskutiert wurden (vgl. hierzu verschiedene Aufsätze von KÜCHLER und LANGEHEINE in der Zeitschrift für Soziologie), in einem Modell vereint werden können;

- es innerhalb des GLIM-Ansatzes auch möglich ist, den Einfluß metrischer Variablen auf eine abhängige qualitative Variable zu untersuchen. Dies ist zwar auch mit den üblichen Methoden multivariater Kreuztabellenanalyse möglich, jedoch liegt diese Erweiterung des Analyserahmens nicht unmittelbar auf der Hand und ist außerdem EDV-technisch - mit Ausnahme von SAS - nicht so leicht durchführbar.

ARMINGER illustriert sein Vorgehen mit einem empirischen Beispiel, in dem unter anderem der Einfluß des Einkommens auf die Erwerbstätigkeit der Frau untersucht wird. Da das Ausgangsmaterial für diese Auswertungen aus kreuztabellierten Daten besteht, müssen durch die Aggregation Informationsverluste bei den metrischen Variablen in Kauf genommen werden. Es kann lediglich der Einfluß des Durchschnittseinkommens des Ehemannes in den durch die unabhängigen (qualitativen) Merkmale definierten Subpopulationen überprüft werden (vgl. Tabelle 1 in ARMINGER, 1983:50). Da im Mikrozensus das Einkommen nach Einkommensklassen erhoben wurde, handelt es sich bei der Variable Z noch nicht einmal um die genauen Durchschnittswerte pro Subpopulation.

Ich möchte in diesem Beitrag, der mehr als ein Nachtrag zu ARMINGERs oben erwähnten Beitrag in der Zeitschrift für Soziologie gedacht ist, zeigen, daß die von ARMINGER berechneten logistischen Modelle ebenso mit Individualdaten untersucht werden können. Auf den formalstatistischen Hintergrund des GLIM-Ansatzes möchte ich dabei nur soweit eingehen, als es für das anwendungsbezogene Verständnis der folgenden Auswertungen notwendig ist. Alle weiteren Details finden sich in ARMINGERs Aufsatz (1983). Diejenigen Leser, die sich intensiver mit der Theorie und Praxis verallgemeinerter linearer Modelle auseinandersetzen wollen, finden dazu weitere Hinweise in dem englischsprachigen Lehrbuch von McCULLAGH & NELDER (1983) sowie in zwei deutschsprachigen Vorlesungsskripten (ARMINGER, 1982; ANDRESS, 1983).

Die folgende Ableitung des logistischen Regressionsmodells für Individualdaten erfordert ebenfalls keine weiteren mathematischen Kenntnisse, sondern erfolgt in Analogie zu bekannten Verfahren multivariater Kreuztabellenanalyse. Das Programmpaket GLIM 3 bildet praktisch nur den didaktischen Rahmen, innerhalb dessen diese Verallgemeinerung unmittelbar evident wird, denn natürlich sind logistische Regressionsmodelle für Individualdaten auch mit anderen Programmpaketen überprüfbar (z.B. Prozedur PLR in BMDP oder Prozedur FUNCAT in SAS).

Dieser allgemeine Rahmen macht es darüber hinaus möglich, strukturelle Ähnlichkeiten, aber auch Unterschiede zwischen der Analyse von Individualdaten, wie man sie aus der klassischen Regressionsanalyse kennt, und der Analyse von aggregierten Daten, wie z.B. Kreuztabellen, zu problematisieren. Auf drei Fragen möchte ich dabei näher eingehen:

1. Bei der Suche nach einem passenden Modell beginnt man bei multivariaten Kreuztabellenanalysen oft mit einem Maximalmodell (saturiertes Modell) und versucht, es möglichst weit zu vereinfachen, ohne die Prognosefehler zu groß werden zu lassen. Umgekehrt beginnt man in der klassischen Regressionsanalyse mit einem Minimalmodell, das sukzessive erweitert wird,

bis die Prognosefehler nicht mehr wesentlich verringert werden können. Über das Ende dieses Suchprozesses entscheidet üblicherweise ein Hypothesentest, der je nach Suchstrategie eine unterschiedliche Zielsetzung hat: Im ersten Fall will man signifikante Abweichungen zwischen Modell und Daten vermeiden (Anpassungstest), während im zweiten Fall alle signifikanten Abweichungen berücksichtigt werden sollen (Signifikanztest). Die Frage ist, ob man eigentlich bei beiden Suchstrategien zu identischen Entscheidungen kommt.

2. Unabhängig vom Ausgang dieser Modelltests möchte man die Anpassung des Modells an die Daten durch ein deskriptives Maß beschreiben. Hierzu kann man auch in der Kreuztabellenanalyse Analoga zu den multiplen und partiellen Bestimmtheitsmaßen R^2 und RSQ CHANGE definieren. Die hier vorgestellte Auswertung zeigt jedoch, daß Modelle für Kreuztabellen nach Maßgabe dieser Kriterien einen sehr viel besseren Modellfit aufweisen als entsprechende Modelle für Individualdaten, obwohl die Parameterschätzungen in beiden Fällen identisch sind. Die Frage ist daher, aufgrund welcher Datenbasis (Individual- oder aggregierte Daten) man ein Bestimmtheitsmaß berechnen soll.

3. Schließlich möchte ich eine Thematik aufgreifen, die in jenen Jahren heftig diskutiert wurde, als die Analyse nicht-metrischer Daten in der Soziologie 'modern' wurde. Nach einer langen Zeit der Anpassung der Daten an die Analysemodelle durch Entwicklung exakter Meßinstrumente ergab sich mit der multivariaten Kreuztabellenanalyse die Möglichkeit, statistische Auswertungen vorzunehmen, die dem Meßniveau der Daten angemessen sind. Von der Scheingenauigkeit bisheriger Analysen war die Rede. In der Auswertung von ARMINGER wird Einkommen sowohl als metrische als auch als qualitative Variable betrachtet, wobei sich in der Tat herausstellt, daß die Unterscheidung verschiedener Einkommensklassen die Daten besser erklärt als die Annahme eines linearen Zusammenhangs zwischen Einkommen und Erwerbstätigkeit. Ein ähnliches Resultat zeigt die folgende Auswertung, allerdings nur, wenn man die aggregierten Daten berücksichtigt. Ist die Entscheidung zwischen metrischen und nicht-metrischen Daten also eine Frage der Datenaufbereitung?

2. Replikation der logistischen Modelle von ARMINGER (1983)

Die von mir analysierten Daten stammen aus dem ALLBUS 1980. Sie haben gegenüber den Mikrozensus-Daten von ARMINGER den Vorteil, daß das Einkommen des Ehemanns in exakten DM-Beträgen erhoben wurde. Auswahl der Untersuchungsgruppe und Definition der Variablen entsprechen den Angaben bei ARMINGER. Auf Grund der geringen Fallzahl der Untersuchungsgruppe (N=340) mußten einige Ausprägungen zusammengefaßt werden. Die betrachteten Modellvariablen sind wie folgt definiert:

X: Einkommen des Ehegatten/Partners (Monatsnettoeinkommen, kategorisiert), wobei

X=1: mehr als 2500 DM
X=2: 1801 - 2500 DM
X=3: bis 1800 DM

A: Schulbildung der Frau, wobei

A=1: kein Schulabschluß, Hauptschulabschluß
A=2: Mittlere Reife (Fachschulreife), Fachhochschulreife, Abitur, Hochschulreife

B: Kinder, wobei

B=1: keine Kinder
B=2: mindestens ein Kind, das älter als 5 Jahre ist
B=3: nur Kinder unter 6 Jahren

R: Zahl der nicht erwerbstätigen Frauen in der Kombination (Subpopulation) XAB

N: Zahl der Frauen insgesamt in der Kombination (Subpopulation) XAB: nicht erwerbstätige Frauen, Arbeiterinnen, Angestellte und Beamtinnen

Z: durchschnittliches Monatseinkommen in der Kombination (Subpopulation) XAB.

Die genauen Daten über die Variablenausprägungen und ihre Häufigkeiten sind in Tabelle 3 (Abschnitt 3) dargestellt. An dieser Stelle interessiert zunächst nur eine möglichst optimale Prognose des Erwerbsverhaltens von Frauen. Analog dem Vorgehen bei ARMINGER wurde ein passendes Modell für die Daten gesucht. Dazu betrachtet man die Devianz, die eine Maßzahl für die Abweichung zwischen Modell und Daten ist und der Summe der quadrierten Residuen in klassischen Regressionsanalysen entspricht.[2] Diese Maßzahl ist näherungsweise Chi^2-verteilt und kann daher für Modelltests verwendet werden. Die entsprechenden Zahlen kann man Tabelle 1 entnehmen; auf Grund der geringeren Anzahl von Merkmalsausprägungen bei den betrachteten Variablen sind alle Devianzen und Freiheitsgrade kleiner als bei ARMINGER.

Die Suche nach dem passenden Modell beginnt zunächst mit dem Maximalmodell (Basismodell 1), das in der Tabelle mit BM1 bezeichnet ist und das alle Haupt- und Interaktionseffekte enthält (saturiertes Modell). In dem Maße, in dem man einzelne Effekte wegläßt, nimmt die Devianz zu, bis von einem gewissen Punkt ab die Abweichungen nicht mehr akzeptabel sind. Alle Modelle mit signifikanter Devianz sollen daher auf Grund ihres mangelnden Datenfits zurückgewiesen werden. Bei der Wahl des Signifikanzniveaus sollte jedoch berücksichtigt werden, daß man einen Anpassungstest durchführt und explizit nicht an signifikanten Ergebnissen interessiert ist. Zu niedrige Signifikanzniveaus (z.B. der übliche 5%-Wert) würden Modelle mit hohen Devianzen nicht von der Modellsuche ausschließen. Einige Autoren schlagen daher vor, ein Signifikanzniveau (z.B. 25%) zu wählen, das es erlaubt, schon kleine Abweichungen vom wahren Modell festzustellen (vgl. z.B. FORTHOFER & LEHNEN, 1981). Als Faustregel kann gelten, daß man alle die Modelle als passend akzeptieren kann, bei denen die Devianz kleiner oder gleich der Anzahl der Freiheitsgrade ist. Das einfachste Modell, das danach die Daten noch angemessen beschreibt, ist das Modell mit den Haupteffekten von A und B. Da mich jedoch insbesondere die Variable Einkommen interessiert, betrachte ich das Modell mit allen drei Haupteffekten eingehender (BM3).

Hier zeigt sich, daß die Weglassung jedes einzelnen Haupteffekts eine signifikante Zunahme der Devianz zur Folge hat. Man beachte, daß es sich bei dieser Überprüfung einzelner Effekte wieder um einen üblichen Signifikanztest handelt. Man ist also an signifikanten Ergebnissen interessiert und ein niedriges Signifikanzniveau (z.B. 5%) ist sinnvoll. Ein Vergleich des Haupteffektmodells A+B mit dem Basismodell 3 zeigt, daß die Weglassung der Variablen Einkommen Z bzw. X eine Erhöhung der Devianz um 4.3 bzw. 5.2 Punktwerte zur Folge hat. Bei einem Signifikanzniveau von 5% ist das eine signifikante (bei Z) oder zumindest auffällige (bei X) Verschlechterung des Modells.

Tabelle 1: Logistische Modelle (aggregierte Daten)

Modell		Vergleich mit Basismodell BM	Effekt d. Variablen	Devianz insg.	Df insg.	B	Devianz Vergleich	Df Vergl.	ΔB
BM1:	A + B + Z + A.B + A.Z + B.Z + A.B.Z		-	4.6	6	0.888	-	-	-
BM2:	A + B + Z + A.B + A.Z + B.Z	BM1	A.B.Z	5.4	8	0.868	0.8	2	0.020
	A + B + Z + A.B + A.Z	BM2	B.Z	5.5	10	0.867	0.1	2	0.001
	A + B + Z + A.B + B.Z	BM2	A.Z	5.6	9	0.865	0.2	1	0.003
	A + B + Z + B.Z + A.Z	BM2	A.B	8.4	10	0.796	3.0+	2	0.073
BM3:	A + B + Z	BM2	A.B , A.Z, B.Z	9.4	13	0.772	4.0	5	0.096
	A + B	BM3	Z	13.7	14	0.668	4.3**	1	0.104
	A + Z	BM3	B	28.8**	15	0.300	19.4***	2	0.472
	B + Z	BM3	A	19.4+	14	0.528	10.0***	1	0.244
BM4:	GM		-	41.2***	17	0.000	-	-	-
BM1:	A + B + X + A.B + A.X + B.X + A.B.X		-	0.0	0	1.000	-	-	-
BM2:	A + B + X + A.B + A.X + B.X	BM1	A.B.X	0.9	4	0.977	0.9	4	0.023
	A + B + X + A.B + A.X	BM2	B.X	3.8	8	0.907	2.9	4	0.070
	A + B + X + A.B + B.X	BM2	A.X	1.6	6	0.961	0.7	2	0.016
	A + B + X + A.X + B.X	BM2	A.B	4.9	6	0.881	4.0+	2	0.097
BM3:	A + B + X	BM2	A.B, A.X, B.X	8.5	12	0.795	7.5	8	0.182
	A + B	BM3	X	13.7	14	0.668	5.2*	2	0.127
	A + X	BM3	B	26.5**	14	0.357	18.1***	2	0.438
	B + X	BM3	A	18.1+	13	0.562	9.6***	1	0.233
BM4:	GM		-	41.2***	17	0.000	-	-	-

+ $p < 0.25$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Zusammenfassend läßt sich am Ende dieses Suchprozesses feststellen, daß Global- und Einzeltests unterschiedliche Empfehlungen darüber geben, welches Modell als passend akzeptiert werden kann. Aus inhaltlichen Gründen bleibe ich daher bei meiner Wahl des Modells mit allen drei Haupteffekten. Grundsätzlich wäre aber zu fragen, ob dieses Dilemma etwas mit der Tatsache zu tun hat, daß beiden Suchstrategien unterschiedliche Zielsetzungen zugrundeliegen (Anpassungs- versus Signifikanztest).

Faßt man schließlich die Devianz des Minimalmodells (BM4), das lediglich aus einer Regressionskonstanten besteht, als Maß der gesamten Streuung in den Daten auf, kann man zusätzlich deskriptive Statistiken zur Beschreibung der Modellanpassung definieren.[3] Statt "erklärter Varianz" spricht man jetzt von "erklärter Devianz". Im einzelnen entspricht der Anteil der Devianz eines konkreten GL-Modells an der Gesamtdevianz (gemessen durch BM4) dem Anteil nicht-erklärter Varianz in der Regressionsanalyse. Ich betrachte im folgenden die erklärte Devianz B, die sich durch Subtraktion von 1 ergibt. Analog entspricht der anteilige Devianzzuwachs, der durch Weglassung einer Variablen entsteht, dem Zuwachs an erklärter Varianz, wenn ein Merkmal zusätzlich in eine Regressionsanalyse aufgenommen wird. Diese partiell erklärte Devianz bezeichne ich mit ΔB.

Danach kommen alle drei Haupteffekte jeweils für 10 bis 45% der gesamten Devianz auf. Die Kinderzahl B erklärt dabei am meisten, gefolgt von der Schulbildung A und dem Einkommen X bzw. Z. Diese Reihenfolge entspricht den Ergebnissen ARMINGERs, wie auch der Tatsache, daß das Einkommen, nicht-metrisch aufgefaßt (Variable X), die Daten besser beschreibt als sein metrisches Pendant Z: 12.7% erklärte Devianz bei X gegenüber 10.4% erklärte Devianz bei Z. Insgesamt erfaßt das Haupteffektmodell BM3 77.2% bzw. 79.5% der gesamten Datendevianz.

Auch eine Inspektion der einzelnen Parameterschätzungen ergibt die gleichen inhaltlichen Ergebnisse wie bei ARMINGER (vgl. Tabelle 2). Die untersuchten Frauen sind umso eher nicht erwerbstätig,

- je geringer ihre Schulbildung ist;
- wenn Kinder, insbesondere solche unter 6 Jahren, vorhanden sind;
- je höher das Einkommen des Ehemannes ist.

Tabelle 2: Parameterschätzungen für die Modelle A+B+Z und A+B+X (aggregierte Daten)

estimate	s.e.	Parameter	
-0.6511	0.4574	<GM	
-0.8697	0.2764	A(2)	A + B + Z
0.6776	0.2951	B(2)	Devianz = 9.4
1.6270	0.3877	B(3)	Df = 13
0.0004	0.0002	Z	
0.5839	0.3858	<GM	
-0.8358	0.2709	A(2)	
0.6647	0.2954	B(2)	A + B + X
1.5760	0.3884	B(3)	Devianz = 8.5
-0.2068	0.3292	X(2)	Df = 12
-0.7084	0.3452	X(3)	

Nachdem wir uns vergewissert haben, daß die Daten aus dem ALLBUS 80 die gleichen substantiellen Schlußfolgerungen erlauben wie die von ARMINGER analysierten Daten,[5] können wir uns der methodischen Seite des Problems zuwenden: Wie lassen sich die gleichen Modelle mit Individualdaten so überprüfen, daß Informationsverluste bei den metrischen Variablen vermieden werden können? Dazu möchte ich noch einmal die Dateneingabe bei GLIM diskutieren und mit anderen Programmen zur multivariaten Kreuztabellenanalyse vergleichen. Es zeigt sich, daß diese verallgemeinerte Form der Dateneingabe den Einschluß metrischer erklärender Variablen und die Verwendung von Individualdaten unmittelbar nahelegt.

3. Kreuztabellenanalyse als Analyse aggregierter Individualdaten

Wenn man die Daten aller 340 Frauen anhand der Merkmale X, A und B sortiert und danach alle Personen zusammenfaßt, die sich bezüglich dieser Merkmale nicht mehr unterscheiden, erhält man Tabelle 3. Aus den ursprünglich 340 Individualdaten werden also 18 homogene Subpopulationen generiert. Ihre Anzahl hängt von der Anzahl der Ausprägungen der Gruppierungsmerkmale ab. Sie sind praktisch die Fälle in einer multivariaten Kreuztabellenanalyse. Will man untersuchen, wie sich der Anteil nicht-erwerbstätiger Frauen in diesen Subgruppen verteilt, dann muß man lediglich die Anzahl nicht-erwerbstätiger Frauen R pro Subpopulation sowie deren Gesamthäufigkeit N auszählen. Diese Häufigkeiten gibt man in ein entsprechendes Auswertungsprogramm[6] ein und kann dann verschiedene lineare, logistische oder log-lineare Modelle testen. Das Programm wird dazu die Charakteristika der Subpopulation (gemessen durch die Merkmale X, A und B) angemessen berücksichtigen und intern die Designmatrix (Matrix der erklärenden Variablen) erzeugen. Mit dieser Aggregation gehen keine Informationsverluste einher, solange alle Merkmale X, A, B und die Variable Erwerbstätigkeit nicht-metrisch sind. Im Gegenteil, Tabelle 3 ist in diesem Fall eine besonders effiziente Reproduktion der ursprünglich sehr viel umfangreicheren Datenbasis.

Tabelle 3: Erwerbstätigkeit verheirateter Frauen im ALLBUS 1980

N	R	X	A	B	Z
4	2	1	1	1	3625.00
26	20	1	1	2	3169.23
8	7	1	1	3	2862.50
6	4	1	2	1	3425.00
22	15	1	2	2	3336.36
8	5	1	2	3	3337.50
15	10	2	1	1	2086.67
63	43	2	1	2	2129.05
24	23	2	1	3	2152.08
8	3	2	2	1	2187.50
22	13	2	2	2	2235.46
11	8	2	2	3	2159.09
25	12	3	1	1	1464.00
50	32	3	1	2	1584.72
25	21	3	1	3	1379.96
8	1	3	2	1	1477.50
10	5	3	2	2	1601.00
3	1	3	2	3	1433.33

Bei der Berücksichtigung metrischer Variablen (z.B. Einkommen) ergibt sich ein doppeltes Problem: Erstens muß man Informationsverluste in Kauf nehmen, wenn man auf der Ebene der Subpopulationen arbeitet, und zweitens können die Ausprägungen dieses erklärenden Merkmals nicht ohne weiteres berücksichtigt werden, da die Designmatrix von gängigen Programmen zunächst nach bestimmten Standardvorgaben erzeugt wird.

Bei der Erzeugung geht man von einem vollständigen faktoriellen Design im Sinne der Varianzanalyse aus, d.h. jede Ausprägungskombination der unabhängigen Merkmale kommt gleich häufig vor, nämlich genau einmal. Kovariate, d.h. metrische Merkmale, sind zunächst ausgeschlossen. Diese Defaults kann man umgehen, indem man die Designmatrix explizit per Hand eingibt, doch ist das nicht bei allen in Anmerkung 6 genannten Programmen möglich. Die geschilderte Eingabe wird durch Tabelle 3 beschrieben. Bei der Aggregation wurden nicht nur die Häufigkeiten N und R, sondern auch das Durchschnittseinkommen Z berechnet. Das Problem ist nun, wie man die 18 verschiedenen Einkommenswerte Z, oder besser noch alle vorhandenen individuellen Einkommen, bei einer Analyse der Erwerbsbeteiligung von Frauen berücksichtigen kann.

Die Antwort ergibt sich unmittelbar, wenn man die Dateneingabe in das Programmpaket GLIM berücksichtigt: GLIM verarbeitet nämlich die gesamte Datenmatrix in exakt der Form, in der sie in Tabelle 3 abgebildet ist. Das schließt nicht aus, daß man, wie in anderen Programmen auch, die Häufigkeiten R und N eingibt. Auch hier kann man die Designmatrix programmintern erstellen lassen. Wenn, wie in unserem Fall, alle Ausprägungen von X, A und B vorkommen, dann kann man diese Variablen mit wenigen GLIM-Befehlen durch das Programm erzeugen lassen. Darüberhinaus hat die Dateneingabe in Form einer Datenmatrix noch eine Reihe weiterer Vorteile (z.B. bei der Behandlung fehlender Werte in Kreuztabellen), auf die ich an dieser Stelle nicht eingehen möchte (vgl. jedoch ANDRESS, 1983:238f.). Bei der Matrixeingabe spielt es überhaupt keine Rolle, ob die Datenmatrix nun aus 18 Subpopulationen oder aus 340 Individuen besteht. Entscheidend ist allein, welche Verteilungsannahmen man über das abhängige Merkmal macht und ob man ein lineares, logistisches oder log-lineares Regressionsmodell zugrundelegen will. Die wesentlichen Annahmen dafür sind bei ARMINGER (s.o.) dargestellt.

Hat man alle Daten für die 18 Subpopulationen an das Programm übermittelt und als abhängiges Merkmal den Anteil R/N gewählt, dann spricht nichts dagegen, den Einfluß aller möglichen Kombinationen der erklärenden Merkmale, also auch den der metrischen Variablen Z zu überprüfen; und zwar genau in der gleichen Weise, wie man mit denselben Daten eine Regressionsanalyse für ein metrisches abhängiges Merkmal rechnen würde. Die verallgemeinerte Form der Dateneingabe legt es also nahe, bei Kreuztabellenanalysen die Beschränkungen bezüglich des Meßniveaus der erklärenden Variablen (keine Kovariate) aufzuheben.

Auch die Analyse von Individualdaten ist unmittelbar evident. Dazu braucht man sich nur einmal zu überlegen, was passiert, wenn jede Subpopulation nur jeweils eine Person umfaßt. Diese "Subpopulationen" sind praktisch Individualdaten. Mit anderen Worten: Da die oben beschriebenen Daten den sonst üblichen Rohdaten strukturell ähnlich sind, liegt eine Verallgemeinerung der im vorigen Abschnitt berechneten Modelle auf Individualdaten auf der Hand. Die Modellstruktur (binomialverteilte Zielvariable, logistisches Regressionsmodell) bleibt gleich, während sich lediglich die Form der Datenaufbereitung ändert.

Insgesamt gesehen verschwinden also die Grenzen zwischen Kreuztabellenanalyse und der Analyse von Individualdaten. Was verbleibt, sind die Modelle, während die Form der Datenaufbereitung eher nebensächlich ist. Der Begriff "Verfahren multivariater Kreuztabellenanalyse" suggeriert ein eigenständiges Methodengebiet, das sich bei näherem Hinschauen als Anwendung von Methoden entpuppt, die gleichermaßen auf andere Datenformen angewendet werden können.

4. Analyse qualitativer Variablen mit Individualdaten ist nichts Neues

Um diese etwas provokante Behauptung zu belegen, erinnern wir und noch einmal an das vorige Gedankenexperiment. Je nachdem, wie differenziert man die obigen Subpopulationen definiert oder wie umfangreich die Stichprobe ist, sind die Besetzungsfehler N der Subpopulationen mehr oder weniger groß. Im Extremfall besteht eine Subpopulation nur aus einer Person. Für die Definition der Modellstruktur sollte es daher keinen Unterschied machen, ob die Häufigkeiten R und N gleich 1 oder größer sind, ob man also Individual- oder aggregierte Daten betrachtet. Konsequenzen hat dies jedoch für die statistische Schätzung der Parameter; eine Beschränkung auf logistische Modelle ist deshalb angezeigt.

Praktisch geht man dabei folgendermaßen vor: Jede Subpopulation besteht jetzt aus einer Person, die Gesamthäufigkeit n_i beträgt also 1. Erneut muß man sich überlegen, wieviele Personen davon nicht erwerbstätig sind. Das kann entweder eine ($y_i=1$) oder gar keine Frau ($y_i=0$) sein. y ist praktisch eine Dummy-Variable, die angibt, ob eine Frau erwerbstätig ($y_i=0$) oder nicht erwerbstätig ($y_i=1$) ist. Diese Dummy-Variable ist das abhängige Merkmal in unserem logistischen Regressionsmodell mit Individualdaten. Sie ist ebenfalls binomialverteilt, denn es gibt zwei mögliche Realisationen dieser Zufallsvariablen bei insgesamt einer Ziehung pro Kombination der unabhängigen Merkmale. Die Wahrscheinlichkeit π_i, "nicht erwerbstätig zu sein", also eine Person mit $y_i=1$ zu beobachten, wird wiederum durch die unabhängigen Merkmale X, A, B und Z erklärt. Die Likelihood-Funktion für die gesamte Stichprobe lautet:

(1) $$L(\underline{\beta}) = \prod_{i=1}^{n} \pi_i(\underline{\beta})^{y_i} (1-\pi_i(\underline{\beta}))^{1-y_i} \quad \text{mit}$$

$$y_1 = \begin{cases} 1 & \text{nicht erwerbstätig} \\ 0 & \text{erwerbstätig} \end{cases}$$ $i = 1,\dots n$ Beobachtungen des abhängigen Merkmals

$$\pi_i(\underline{\beta}) = \exp(\eta_i)/(1+\exp(\eta_i))$$ inverse Verbindungsfunktion Logit

$$\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j$$ linearer Prädiktor mit Parametervektor $\underline{\beta} = (\beta_j)$ $j = 1,\dots p$ und unabhängigen Merkmalen x_{ij}

ML-Schätzer der unbekannten Parameter β lassen sich mit GLIM oder einem entsprechenden anderen Programm berechnen, indem man y als abhängiges Merkmal deklariert und die Binomialverteilung sowie die Verbindungsfunktion Logit verwendet. N hat jetzt für alle Beobachtungen den Wert 1.

Wie aus Tabelle 4 zu erkennen ist, ergeben sich (bis auf Rundungsfehler) die gleichen Schätzwerte für das Haupteffektmodell A+B+X. Dieses Ergebnis war zu erwarten, denn die Daten enthalten keine neuen Informationen gegenüber den aggregierten Daten in Tabelle 3. Lediglich die Gesamtdevianz und die Freiheitsgrade haben sich auf Grund der erhöhten Fallzahl vergrößert, doch darauf möchte ich im nächsten Abschnitt eingehen.

Tabelle 4: Parameterschätzungen für die Modelle A+B+Z und A+B+X (Individualdaten)

estimate	s.e.	Parameter	
-0.8988	0.4383	<GM	
-0.9372	0.2760	A(2)	A + B + Z
0.6574	0.2962	B(2)	Devianz = 398.6
1.6420	0.3897	B(3)	Df = 335
0.0005	0.0002	Z	
0.5838	0.3852	<GM	
-0.8357	0.2705	A(2)	
0.6647	0.2951	B(2)	A + B + X
1.5760	0.3879	B(3)	Devianz = 402.4
-0.2068	0.3285	X(2)	Df = 334
-0.7083	0.3445	X(3)	

Andere Resultate ergeben sich allerdings bei dem Haupteffektmodell A+B+Z, in dem Einkommen als metrische Variable betrachtet wird. Durch den Übergang auf die Ebene der Individualdaten ist es jetzt möglich, die Originalwerte von Z statt der Subpopulationsdurchschnitte zu verwenden und somit den Informationsverlust zu umgehen, der mit der Aggregation verbunden ist. Inhaltlich ergeben sich zwar keine neuen Einsichten, die Wahrscheinlichkeit der Nicht-Erwerbstätigkeit wächst weiterhin mit dem Einkommen, jedoch hat sich dieser Einfluß verstärkt und die Signifikanz des Parameters ist größer geworden. Der entsprechende t-Wert beträgt jetzt t=2.88.

Ähnliche Überlegungen lassen sich auch für log-lineare Modelle anstellen: Auch hier läßt sich zeigen, daß die gleichen Modelle genauso gut mit Individualdaten geschätzt werden können. Auf Grund der geringen Fallzahl meiner Stichprobe habe ich jedoch davon abgesehen, ARMINGERs Modell für verschiedene Arten von Erwerbstätigkeit (1983:61ff.) mit meinen Daten zu replizieren.

Fassen wir einmal zusammen, was wir bis jetzt wissen: Methoden, die wir bisher aus der Analyse multivariater Kreuztabellen kennen (logistische und log-lineare Modelle), lassen sich genauso gut auf Individualdaten anwenden.

Wenn alle Daten nur aus qualitativen (nicht-metrischen, diskreten) Merkmalen bestehen, sind Kreuztabellen bloß eine besonders effiziente Form der Datenaufbereitung, denn angesichts der begrenzten Zahl von Merkmalsausprägungen werden mehrere Analyseeinheiten identische Werte bei unabhängigen und abhängigen Variablen aufweisen. Auswertungen mit den Originaldaten wie mit den aggregierten Daten liefern identische Parameterschätzungen. Anders hingegen bei metrischen Variablen: Hier sind mit der Aggregation immer Informationsverluste verbunden, sodaß es sinnvoller ist, bei der Betrachtung metrischer Variablen die Original- oder Individualdaten zu verwenden.

Nun bestand die bisherige Auswertung nicht nur aus der Schätzung von Variableneffekten, es wurden auch verschiedene Erklärungsmodelle für die Daten gegeneinander getestet (vgl. Tabelle 1). Nachdem gezeigt wurde, daß Analysen mit Individualdaten identische bzw. präzisere Parameterschätzungen ergeben, stellt sich jetzt die spannende Frage, ob man eigentlich das gleiche Haupteffektmodell als passende Beschreibung der Daten gewählt hätte, wenn man gleich mit der Auswertung von Individualdaten begonnen hätte. Dieser Frage will ich im letzten Abschnitt nachgehen, wo ich die Ergebnisse der bisherigen Auswertung problematisiere.

5. Diskussion der Auswertung mit Individualdaten

Die Ergebnisse der verschiedenen Modelltests sind analog dem bisherigen Vorgehen in Tabelle 5 dargestellt. Auf Grund der höheren Fallzahl (340 Frauen statt 18 Subpopulationen) sind alle Gesamtdevianzen und Freiheitsgrade höher. Wenn man allerdings Modelle mit dem entsprechenden Basismodell vergleicht, um den Effekt einer Variablen zu evaluieren, dann sind die Devianz- bzw. Freiheitsgraddifferenzen wieder gleich. Mit anderen Worten: Die Devianz hat sich zwar absolut erhöht, aber die Relationen zwischen den Modellen bleiben erhalten. Ausgenommen sind Modelle mit Z als unabhängiger Variablen: Hier werden jetzt andere Daten, nämlich die Originalwerte statt der Subpopulationsdurchschnitte, berücksichtigt.

Nach den oben diskutierten Kriterien eines Anpassungstests beschreibt keines der Modelle die Daten adäquat. Alle Devianzen sind weit größer als die jeweiligen Freiheitsgrade. Darüber hinaus erklärt keines der Modelle mehr als 10% der Gesamtdevianz gegenüber fast 80% erklärter Devianz bei den oben

Tabelle 5: Logistische Modelle (Individualdaten)

Modell		Vergleich mit Basismodell BM	Effekt d. Variablen	Devianz insg.	Df insg.	B	Devianz Vergleich	Df Vergl.	ΔB
BM1:	A + B + Z + A.B + A.Z + B.Z + A.B.Z		-	393.1***	328	0.097	-	-	-
BM2:	A + B + Z + A.B + A.Z + B.Z	BM1	A.B.Z	393.3***	330	0.096	0.2	2	0.000
	A + B + Z + A.B + A.Z	BM2	B.Z	394.6***	332	0.093	1.3	2	0.003
	A + B + Z + A.B + B.Z	BM2	A.Z	393.3***	331	0.096	0.0	1	0.000
	A + B + Z + B.Z + A.Z	BM2	A.B	398.3***	332	0.085	5.0*	2	0.012
BM3:	A + B + Z	BM2	A.B , A.Z, B.Z	398.6***	335	0.084	5.3	5	0.012
	A + B	BM3	Z	407.6***	336	0.063	9.0***	1	0.021
	A + Z	BM3	B	418.2***	337	0.039	19.6***	2	0.045
	B + Z	BM3	A	410.3***	336	0.057	11.7***	1	0.027
BM4:	GM		-	435.1***	339	0.000	-	-	-
BM1:	A + B + X + A.B + A.X + B.X + A.B.X		-	393.9***	322	0.095	-	-	-
BM2:	A + B + X + A.B + A.X + B.X	BM1	A.B.X	394.9***	326	0.093	0.9	4	0.002
	A + B + X + A.B + A.X	BM2	B.X	397.7***	330	0.086	2.9	4	0.007
	A + B + X + A.B + B.X	BM2	A.X	395.5***	328	0.091	0.7	2	0.002
	A + B + X + A.X + B.X	BM2	A.B	398.8***	328	0.083	4.0+	2	0.009
BM3:	A + B + X	BM2	A.B, A.X, B.X	402.4***	334	0.075	7.5	8	0.017
	A + B	BM3	X	407.6***	336	0.063	5.2*	2	0.012
	A + X	BM3	B	420.4***	336	0.034	18.1***	2	0.042
	B + X	BM3	A	412.0***	335	0.053	9.6***	1	0.022
BM4:	GM		-	435.1***	339	0.000	-	-	-

+ $p < 0.25$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

diskutierten Haupteffektmodellen für aggregierte Daten. Sind die Modelle jetzt auf einmal soviel schlechter, obwohl sie identische bzw. sogar präzisere Parameterschätzungen liefern? Natürlich nicht: Die abweichenden Ergebnisse sind allein Resultat der unterschiedlichen Fallzahl, die bei aggregierten Daten immer von der Anzahl der Ausprägungskombinationen (=Anzahl der Subpopulationen) abhängt. Die Devianz des Minimalmodells BM4 in Tabelle 1 ist also gar nicht die Gesamtdevianz der Daten, sondern lediglich eine Teildevianz, nachdem schon 18 Subpopulationen zusammengefaßt wurden. Allein die Devianz desselben Minimalmodells für Individualdaten (435.1, vgl. Tabelle 5) gibt Auskunft über die Gesamtstreuung der Daten. Sie ist daher die richtige Bezugsgröße für alle multiplen und partiellen Bestimmtheitsmaße, die von der Anzahl der erklärenden Merkmale und deren Ausprägungen unabhängig sind. Bei einer Analyse der erklärten Devianz sollte man daher immer die entsprechenden B- und ΔB-Koeffizienten für Individualdaten berücksichtigen. Bei aggregierten Daten (Kreuztabellen) hingegen suggerieren diese Statistiken einen sehr viel besseren Modellfit, als er in Wahrheit vorliegt.

Allerdings hilft die Richtigstellung bei der Suche nach einem passenden Modell auch nicht viel weiter, denn sie liefert keine klaren Kriterien für die Modellwahl. Von welchem Ausmaß erklärter Devianz ab sollte man etwa ein Modell als passend akzeptieren? Eine weitere Auswahlstrategie wäre, keinen Effekt zu vernachlässigen, der für sich genommen einen signifikanten Einfluß hat. Mit anderen Worten: Man begründet die Modellwahl mit den Signifikanztests, die den Effekt einzelner Variablen überprüfen (vgl. Spalten 6,7 in Tabelle 5). Danach würde man auch in diesem Fall das Basismodell 3 als passend akzeptieren, denn alle drei Haupteffekte sind signifikant. Häufig versagt aber auch dieses Kriterium (vgl. Tabelle 4 bei ARMINGER, 1983:54, in der alle Variablen für sich genommen einen signifikanten Einfluß zeigen).

Was verbleibt, ist die erklärte Devianz jeder Variablen (vgl. Spalte 8 in Tabelle 5). Die höchsten Werte erzielen die drei Haupteffekte, insgesamt gesehen sind die Werte jedoch im Vergleich zu Tabelle 1 sehr viel niedriger (aus den oben genannten Gründen). Erneut entsteht das Problem, wo man die Grenze zwischen relevanten und irrelevanten Einflüssen ziehen soll. In Tabelle 1 würde die Wahl sehr viel einfacher fallen, da die Unterschiede sehr viel krasser sind. In Tabelle 5 hingegen ist das Niveau insgesamt sehr viel geringer, und es ist daher sehr viel schwerer, eine eindeutige Trennungslinie zu ziehen.

Noch etwas fällt auf, wenn man sich die ΔB-Koeffizienten in beiden Tabellen anschaut. Danach erklärt bei den aggregierten Daten das Einkommen als qualitative Variable sehr viel mehr als die metrische Variable Z (12.7% erklärte Devianz bei X versus 10.4% bei Z). Bei den Individualdaten stellt sich die Situation genau andersherum dar (1.2% erklärte Devianz bei X versus 2.1% bei Z). Nach diesem Kriterium hat der Haupteffekt von X genausoviel Einfluß wie der Interaktionseffekt von A und B.

Nachdem wir zunächst glücklich festgestellt haben, daß Logitanalysen mit Individualdaten möglich sind, sind wir nun wieder gründlich verwirrt, da diese Auswertungen zwar identische oder sogar präzisere Parameterschätzungen liefern, wir aber bei der Suche nach einem passenden Modell möglicherweise zu ganz anderen Resultaten kommen.

6. Schlußfolgerungen

Aus den im letzten Abschnitt dargestellten Ergebnissen ergeben sich die folgenden Schlußfolgerungen:

1. Zunächst einmal muß festgestellt werden, daß bei der Modellsuche und Überprüfung der einzelnen Effekte zwei ganz unterschiedliche Teststrategien eingesetzt werden. Im ersten Fall (Globaltest) verwendet man einen Anpassungstest, während bei der Evaluierung einzelner Variablen übliche Signifikanztests eingesetzt werden. Bekanntermaßen sind alle Tests bei großen Fallzahlen anfällig: Auch der kleinste Unterschied wird signifikant bzw. die kleinste Abweichung hat eine mangelnde Anpassung zur Folge. Große Fallzahlen haben daher beim GLIM-Ansatz zur Folge, daß keines der Modelle (außer dem saturierten oder einigen anderen sehr komplexen Modellen) die Daten adäquat beschreibt, während jeder Variableneffekt für sich genommen signifikant ist.

2. Die Strategie eines Anpassungstests ist an sich eine sehr plausible Vorgehensweise: Man beginnt zunächst mit einem Modell (das saturierte Modell),

das die Daten vollständig erklärt. Dieses Modell ist praktisch eine "Nacherzählung" der Daten ohne Herausstellung der wesentlichen Datenstrukturen. Dann reduziert man schrittweise den Komplexitätsgrad und überlegt sich, wie weit man den Informationsgehalt des Modells reduzieren kann, ohne wesentliche Details der Daten zu vernachlässigen. Das passende Modell schließlich stellt einen optimalen Kompromiß zwischen Informationsreduktion und Detaillierungsgrad dar. Kriterium dafür ist die Abweichung (Devianz) der Modellprognosen von den empirischen Werten.

3. Allerdings ist klar, daß ein solches saturiertes Modell praktisch nur für Kreuztabellen oder aggregierte Daten verwendet werden kann. Bei Individualdaten hätte dieses saturierte Modell so viele Parameter, daß es die Kapazität jedes Programmpaketes sprengen würde. Die oben diskutierten Anpassungstests stehen also in der Tradition der Kreuztabellenanalyse.

4. Erinnern wir uns aber einmal, wie die entsprechenden Globaltests bei klassischen Regressionsanalysen verlaufen. Man formuliert ein Minimalmodell,[7] das nur aus der Regressionskonstanten besteht und überprüft mit Hilfe eines multiplen F-Tests, ob die Hinzufügung weiterer unabhängiger Variablen eine signifikante Modellverbesserung ergibt. Während also bei der Kreuztabellenanalyse die Suche nach dem passenden Modell vom komplexesten zum einfachsten Modell erfolgt, ist es bei der Regressionsanalyse genau umgekehrt. Diese umgekehrte Suchstrategie ist aber auch bei GL-Modellen einsetzbar. Auch hier kann man die Modellverbesserung (relativ zu einem Minimalmodell) überprüfen, statt der Frage nachzugehen, ob ein gegebenes Modell die Daten angemessen beschreibt. Diese umgekehrte Vorgehensweise hätte zwei Vorteile:

a) Die Überprüfung der Modellanpassung sowie der einzelnen Variableneinflüsse würde beide Male mit Hilfe eines Signifikanztests geschehen. Damit wären solche Ergebnisse ausgeschlossen, wie sie oben beschrieben worden sind: Jede Variable ist signifikant, aber kein Modell paßt. Bei großen Fallzahlen wäre immer noch jede Variable signifikant, aber auch jedes Modell ergäbe eine signifikante Verbesserung und würde damit akzeptiert. Wahrscheinlich ist, daß ab einem gewissen Komplexitätsgrad des Modells weitere erklärende Variablen keine signifikanten Modellverbesserungen zur Folge haben. Man wird die Modellsuche also irgendwann abbrechen, bevor man zum saturierten Modell kommt.

b) Grundlage solcher Tests wäre immer die Differenz der Devianzen des untersuchten und des Minimalmodells, niemals der absolute Betrag der Devianz, welcher, wie wir oben gesehen haben, immer von der Fallzahl abhängt.[8] Würde man also die Modellverbesserung testen, gäbe es keine Unterschiede zwischen der Analyse von aggregierten und der von Individualdaten.

5. Generell sind alle Tests nicht besonders aussagekräftig, wenn man große Stichproben betrachtet. Aus dieser Notlage heraus verwendet man deskriptive Maße des Modellfits (multiple und partielle Bestimmtheitsmaße), und verschiedene Autoren haben gezeigt, daß solche Maße auch für Analyseverfahren berechnet werden können, die nicht zum klassischen linearen Modell (Regressions-, Varianz- und Kovarianzanalyse) gehören. Für GL-Modelle konnte gezeigt werden, daß Maßzahlen erklärter Devianz (B- und ΔB-Koeffizient) nur dann von der Anzahl der erklärenden Merkmale und ihren Ausprägungen unabhängig sind, wenn sie auf der Basis von Individualdaten berechnet werden. Dieses Ergebnis zeigt erneut die Notwendigkeit von Auswertungen mit Individualdaten. Darüberhinaus sind bei Berücksichtigung mehrerer Variablen präzisere Schätzungen möglich, da keine Informationsverluste durch Aggregation entstehen.

6. Schließlich relativiert mein Vergleich des Einflusses der Variablen Z bzw. X die frühere Debatte über die Scheingenauigkeit metrischer Merkmale. Man kann einwenden, daß dieses Ergebnis möglicherweise ein Einzelfall ist, doch haben ähnliche Auswertungen mit anderen Daten (vgl. ANDRESS, 1983) ebenfalls gezeigt, daß metrische Merkmale gegenüber ihren qualitativen Pendants bei Individualdaten immer bessere Prognosen liefern. Umgekehrt ergeben qualitative Merkmale bei aggregierten Daten immer bessere Prognosen. Hier sind weitere Forschungen nötig, um dieses Datenproblem zu klären. Das schließt natürlich nicht aus, daß bei jeder metrischen Variablen geprüft werden muß, ob die Annahme eines linearen Zusammenhangs gerechtfertigt ist.

Dieser Bericht wurde von Hans-Jürgen Andreß (Universität Bielefeld) verfaßt, der an mehreren ZUMA-Arbeitstagungen als Referent mitgewirkt hat.

Anmerkungen

1) General linear modeling. Im folgenden verwende ich auch häufig die Bezeichnung GL-Modell für ein verallgemeinertes lineares Modell.
2) In der Tat läßt sich zeigen, daß die Devianz in einem GLIM-Modell mit einer normalverteilten Zielvariablen exakt der Summe der quadrierten Residuen entspricht.
3) In Abschnitt 5 werde ich zeigen, daß dieses Minimalmodell nicht die adäquate Bezugsgröße ist, da seine Devianz von der Anzahl der erklärenden Merkmale und ihren Ausprägungen abhängt. Alle Anteile erklärter Devianz sollten vielmehr auf die Devianz des entsprechenden Minimalmodells für Individualdaten bezogen werden.
4) Nicht alle Ausprägungen des Merkmals X sind von Null verschieden. Der t-Wert für die Variable Z ist gerade signifikant (t=2.03).
5) Auf Grund der höheren Fallzahl sind bei ARMINGER noch zwei Interaktionseffekte (A.B und B.Z bzw. B.X) signifikant.
6) Z.B. ECTA oder P3F im BMDP (ML-Schätzung) sowie NONMET oder GENCAT (WLS-Schätzung). Eingabedaten sind genauer gesagt die Häufigkeiten R bzw. (N-R).
7) Ein sogenanntes "no information"-Modell: Mangels besserer Informationen nimmt man den Durchschnitt des abhängigen Merkmals als beste Prognose. Es ließe sich aber auch jedes andere Modell (z.B. das Modell, das die allgemein anerkannten Einflußfaktoren kontrolliert) als Minimalmodell verwenden. Faktisch ist der F-Test ein Test auf Weglassung der in Frage stehenden Variablen. Das ändert jedoch nichts an der prinzipiellen Suchstrategie, die vom einfachsten zum komplexesten Modell verläuft.
8) Wir erinnern uns: Die Signifikanztests der einzelnen Variableneffekte lieferten identische Ergebnisse (bis auf die Variable Z), egal, ob wir aggregierte oder Individualdaten betrachtet haben. Sie verwendeten lediglich Devianzdifferenzen. Anders hingegen bei den Globaltests, die auf der absoluten Devianz beruhten: Bei der Verwendung von Individualdaten paßte kein Modell auf die Daten.

Literatur

ANDRESS, H.J. Verallgemeinerte lineare Modelle - Eine Einführung in das Programm GLIM. Bielefeld: Unveröffentlichtes Vorlesungsmanuskript, 1983.

ARMINGER, G. Klassische Anwendungen verallgemeinerter linearer Modelle in der empirischen Sozialforschung. In: Verallgemeinerte lineare Modelle in der empirischen Sozialforschung. NONMET/GLIM-Workshop 16.-20.11.81. ZUMA-Arbeitsbereicht, Nr. 1982/03.

ARMINGER, G. Multivariate Analyse von qualitativen abhängigen Variablen mit verallgemeinerten linearen Modellen. Zeitschrift für Soziologie, 1, 1983, 49-64.

FORTHOFER, R.N. & LEHNEN, R.G. Public program analysis. Belmont, CA: Wadsworth, 1981.

McCULLAGH, P. & NELDER, J.A. Generalized linear models. London: Chapman & Hall, 1983.

NELDER, J.A. & WEDDERBURN, R.W.M. General linear models. Journal of the Royal Statistical Society, Series A, 135, 1972, 370-384.