

Diskussion

Kommentar zu Rolf Langeheine, „Multivariate Analyse nominalskalierter Daten via Goodmans Modell: Sehr wohl eine Alternative“ (ZfS Oktober 1979)**Manfred Küchler**Universität Frankfurt, Fachbereich Gesellschaftswissenschaften
Senckenberganlage 15, D-6000 Frankfurt am Main

Es ist zu begrüßen, daß auch in der Bundesrepublik eine Diskussion um neuere Verfahren der Datenanalyse in Gang kommt, nicht zuletzt gefördert durch die redaktionelle Politik der Zeitschrift für Soziologie. Bedauerlich an Langeheines Aufsatz ist jedoch, daß er nicht den jüngsten Diskussionsstand aufnimmt, sondern sich über weite Strecken an Positionen abarbeitet, die längst revidiert sind, und zwar nicht nur in Form von ‚persönlichen Mitteilungen‘, sondern allgemein zugänglich (Küchler 1978, 1979). Warum sich ein Autor trotz besseren Wissens beharrlich auf überholte Diskussionsbeiträge bezieht, und warum eine angesehene Zeitschrift dies in ihrem Begutachtungsverfahren nicht beanstandet, mag ein interessantes wissenschaftssoziologisches Problem sein, doch diesem will ich hier nicht nachgehen.

Im folgenden soll lediglich versucht werden, thesenartig darzustellen, welche Punkte tatsächlich kontrovers sind. Für eine umfassendere Darstellung der Gesamtproblematik sei auf mein Lehrbuch (Küchler 1979a), für eine elementare Einführung auf ein Vorlesungskript (Küchler 1979b) verwiesen.

1. Harders DO-Modell ist ein relativ schlichtes Modell, das zweifelsohne beträchtliche statistische Defizite aufweist, die durch die zunächst von mir vorgeschlagene Verbindung mit Teilen von Goodmans Ecta (Küchler 1976) nicht optimal behoben werden können.
2. Harders DO-Modell ist ein Spezialfall des allgemeinen GSK-Ansatzes, innerhalb dessen die statistischen Schätz- und Testprobleme adäquat gelöst werden können.
3. Das DIEC-Programm wird deshalb seit Frühjahr 1978 nicht mehr vertrieben, sondern ist durch das DO-Programm ersetzt worden. Die eigenständige Existenz eines DO-Programms ist dadurch gerechtfertigt, daß dieses Modell in ei-

ner ersten Phase der Datenanalyse gute Dienste leisten kann, wenn die Zahl der möglichen Erklärungsmodelle noch sehr groß ist.

4. Für die Detailanalyse, also nach Feststellung, welche Merkmale überhaupt in das endgültige Modell eingehen sollen, ist die Benutzung elaborierterer Ansätze unbedingt notwendig.

5. Der Unterschied zwischen dem GSK-Ansatz und dem von Goodman besteht *nicht* darin, daß einmal mit Anteilswerten, das andere Mal mit log-odds gearbeitet werden muß. Vielmehr bietet der GSK-Ansatz beide Möglichkeiten, neben vielen weiteren Optionen. Seine ganze Anwendungsbreite wird in einem noch in Vorbereitung befindlichen Buch von Lehnen und Forthofer dargestellt.

6. Die Entscheidung für oder gegen ‚log-odds‘ hängt im wesentlichen davon ab, welche ‚Philosophie‘ der Auswerter verfolgt. Liegt der Schwerpunkt der Auswertung darauf, zunächst einmal die Komplexität der Daten präzise, aber auch möglichst anschaulich zu erfassen, dann ist es wichtig, daß die berechneten Koeffizienten für sich verständlich sind. Verfolgt man eher die Strategie des ‚Theorie-Testens‘ – wie etwa Duncan (1979) –, geht man also von einem ex ante postuliertem Modell aus, dann ist es relativ gleichgültig, ob die Koeffizienten am Ende intuitiv plausibel sind. In diesem Fall ist der Vergleich zwischen berechneten und aus dem postulierten Modell abgeleiteten Werten das einzig entscheidende.

7. Es ist kein Vorzug per se, wenn man eine Datenkonstellation z.B. bei Verwendung der log-odds mit weniger Parametern beschreiben kann. Auch bei Verwendung von Anteilswerten läßt sich – auf Kosten der Anschaulichkeit – eine Reduzierung der benötigten Parameter in vielen Fällen erreichen (vgl. etwa Higgins und Koch 1977)

8. Der log-lineare Ansatz bietet gewisse formalstatistische Vorzüge und läßt sich – soweit dies

der mit der Auswertung verbundenen Intention nicht entgegensteht – im GSK-Ansatz ebenfalls verfolgen, ohne daß hierbei wie bei Goodman Inkonsistenzen zwischen der Signifikanzbetrachtung der Effekte und dem Testen analoger Hypothesen auftreten können. Diese Inkonsistenzen lassen sich allerdings durch die Einbeziehung nicht-hierarchischer Hypothesen nach Magidson – die Langeheine dankenswerterweise dem deutschen Publikum näherbringt – mildern.

9. Die Diskussion um die Alternative GSK/ Goodman kann (und wird auch) auf der Ebene formalstatistischer Argumente (generalisierte Chi-Quadrat-Methode versus Maximum-Likelihood-Methode) geführt werden; eine Ebene, die dem Sozialwissenschaftler in der überwiegenden Mehrzahl der Fälle nicht zugänglich ist. Die statistische Validität des GSK-Ansatzes wird jedoch auch von Langeheine nicht bestritten.

10. Insgesamt halte ich also meine These aufrecht, daß dem Goodman-Ansatz Bedeutung im wesentlichen für ‚nicht-metrische Pfadanalysen‘ zukommt, daß aber bei asymmetrischen Fragestellungen (eindeutig festgelegter Zielvariablen) der GSK-Ansatz wegen seiner größeren Flexibilität vorzuziehen ist; auch und gerade wenn das Herz des Forschers an log-odds hängt. Schließlich, und das hätte man mit Recht an

meinen bisher vorgelegten Arbeiten kritisieren können, habe ich es bislang versäumt, auf weitere programmtechnische Umsetzungen des GSK-Ansatzes hinzuweisen. Hier ist insbesondere das von dem ‚geistigen Vater‘ Gary G. Koch mitentwickelte Programm Gencat zu nennen, das nun in einer verbesserten Form als Miscat vorliegt (Stanish/Koch/Landis 1978).

Literatur:

- Duncan, O.D., 1979: Constrained parameters in a model for categorical data. *Sociological Methods & Research* 8: 57–68.
- Higgins, J.E./Koch, G.G., 1977: Variable Selection and Generalized Chi-Square Analysis of Categorical Data applied to a Large Cross-Sectional Occupational Health Survey. *International Statistical Review* 45: 51–62.
- Küchler, M., 1978: Alternativen in der Kreuztabelleanalyse. *Zeitschrift für Soziologie* 7: 347–365.
- Küchler, M., 1979a: Multivariate Analyseverfahren. Stuttgart.
- Küchler, M., 1979b: Einführung in die multivariate Analyse nichtmetrischer Daten (Vorlesungskript). Universität Frankfurt.
- Langeheine, R., 1979: Multivariate Analyse nominalskalierteter Daten via Goodmans Modell: Sehr wohl eine Alternative. *Zeitschrift für Soziologie* 8: 380–390.
- Stanish, W.M./Koch, G.G./Landis, J.R., 1978: A computer program for multivariate ratio analysis (Miscat). *Computer Programs in Biomedicine* 8: 197–206.

Antwort auf Manfred Küchler

Rolf Langeheine

Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel
Olshausenstraße 40–60, D-2300 Kiel 1

Es war einmal ein stolzes Kind namens DO, ein Computerprogramm zur Analyse multivariater Kreuztabellen. DO – oder besser: das hinter ihm stehende Modell – erwies sich allerdings als etwas schwach auf den Beinen. Von einem Verwandten bekam es daher Krücken und einen neuen Namen: DIEC. Die Krücken stellten sich jedoch als Tücken heraus. Sie wurden deshalb durch Anleihen bei einem sehr nahen Verwandten ersetzt. Seitdem heißt DO wieder DO und kann manches, was auch seine größeren Brüder können. Man kann sich viele Vergleiche

vorstellen. Mir gefällt dieser am besten: DO und NONMET II etwa verhalten sich so wie die Rubrik „Weine“ in den Getränkekarten eines norddeutschen und eines süddeutschen Restaurants.

Um es explizit zu sagen: ich werte eine Entwicklung als positiv, in der Schwächen überwunden werden. Da jedoch andere Programme für den GSK-Ansatz dasselbe und viel mehr leisten als DO, plädiere ich für den zweiten Teil von „love it or leave it“. Manchmal muß man halt auch bei Lieblingskindern hart sein. Dies ist zugleich ein Kommentar zu den ersten 4 von Küchler's (1980) 10 Thesen (Anmerkungen) zu meiner Arbeit (Langeheine 1979). Man wird übrigens schnell sehen, daß die Punkte 1 und 2 gar nicht kontrovers sind. Gleichermaßen gilt dies für den 4. Punkt, sofern man das Suchen nach einem sog. besten Modell gegenüber dem