Statistische Modellansätze in der Kontext-Analyse

Wiedenbeck, Michael; Rothe, Günter

Statistische Modellansätze in der Kontext-Analyse

1. Vorbemerkung

Die Kontext-Analyse befaßt sich mit Modellen zur Beschreibung der Verhaltensweisen von Individuen derart, daß nicht nur individuelle Einflußgrößen, sondern darüber hinaus auch sogenannte "Kontextvariablen" berücksichtigt werden, die für alle Individuen einer vorgegebenen Gruppe gleich sind. Die in diesem Zusammenhang zu nennenden Modellansätze sind insbesondere durch das Lehrbuch von Boyd und Iversen (1979) populär geworden. Im vorliegenden Beitrag soll nun versucht werden, für zwei recht elementare, aber dennoch typische Ansätze präzise mathematische Modelle zu formulieren und hiermit einige Schätzverfahren zu vergleichen bzw. Interpretationsaspekte zu diskutieren.

2. Das Standardmodell der Kontext-Analyse

Beide in diesem Beitrag untersuchten Ansätze lassen sich im weiteren Sinne als Lineares Modell mit speziellen Design- und Kovarianzstrukturen auffassen. Im einfachsten Ansatz (vgl. hierzu auch Erbring/Young 1980) ist der Vektor der Responses y_{ij} der Individuen $j=1,...,n_i$ in den Gruppen $i=1,...,m$ die abhängige Variable eines Linearen (Regressions-)Modells, das als unabhängige Einflußgrößen sowohl K verschiedene "individuelle" Variablen als auch entsprechend K verschiedene "Kontextvariablen" besitzt.

Wir betrachten zunächst die Vektoren $x_i^{(\ell)}$, $\ell=1,...,K$, der individuellen unabhängigen Variablen für jede Gruppe $i=1,...,m$

$$(2.1) \qquad x_i^{(\ell)} = \begin{bmatrix} x_{i1}^{(\ell)} \\ \vdots \\ x_{in_i}^{(\ell)} \end{bmatrix}$$

und führen folgende Bezeichnungen ein:

(2.2) $$X_i = \left[x_i^{(1)}, \ldots, x_i^{(K)} \right]$$

sei die $(n_i xK)$-Matrix, deren ℓ-te Spalte der Vektor $x_i^{(\ell)}$ ist. Ferner sei

(2.3) $$x^{(\ell)} = \begin{bmatrix} x_1^{(\ell)} \\ \vdots \\ x_m^{(\ell)} \end{bmatrix}$$

also ein Vektor der Dimension $n = \sum_{i=1}^{m} n_i$ sowie

(2.4) $$X = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}$$

die (nxK)-Matrix, deren ℓ-te Spalte $x^{(\ell)}$ ist.

Die Kontextvariablen werden nun in der Regel folgendermaßen gebildet: Sei zunächst $\bar{x}_i^{(\ell)}$ der Mittelwert von $x_i^{(\ell)}$, also

(2.5) $$\bar{x}_i^{(\ell)} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}^{(\ell)} \quad ,$$

und $\bar{x}_i$ der K-dimensionale Zeilenvektor

(2.6) $$\bar{x}_i = (\bar{x}_i^{(1)}, \ldots, \bar{x}_i^{(K)}) \quad .$$

Sei weiter die $(n_i xK)$-Matrix $\bar{X}_i$ definiert durch

(2.7) $$\bar{X}_i = 1_{n_i} \bar{x}_i \quad ,$$

wobei 1_q den q-dimensionalen Spaltenvektor $(1,1,\ldots,1)'$ bezeichnet sowie

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_m \end{bmatrix} . \tag{2.8}$$

Damit ist $\bar{X}$ eine nxK-Matrix, die aus der nxK-Matrix X dadurch hervorgeht, daß in jeder Spalte von X die Komponenten $x_{ij}^{(\ell)}$ durch die Mittelwerte $\bar{x}_i^{(\ell)}$ in den jeweiligen Gruppen ersetzt werden.

Nach diesen Vorbereitungen läßt sich das Standardmodell der Kontextanalyse

$$Y_{ij} = \alpha + \Sigma_{\ell=1}^{K} x_{ij}^{(\ell)} \beta_\ell^{(1)} + \Sigma_{\ell=1}^{K} \bar{x}_i^{(\ell)} \beta_\ell^{(2)} + e_{ij} \tag{2.9}$$

schreiben als

$$Y = \left[1_n, X, \bar{X} \right] \begin{bmatrix} \alpha \\ \beta^{(1)} \\ \beta^{(2)} \end{bmatrix} + e , \tag{2.10}$$

wobei wir wie üblich annehmen, daß für den Erwartungswert und die Kovarianz des Fehlerterms e gilt:

$$E(e)=0 \text{ und } Kov(e)= \sigma^2 I_n; \tag{2.11}$$

hierbei bezeichnet I_n die n-dimensionale Einheitsmatrix.

Das Modell läßt sich nun unter Verwendung einer Aggregationsmatrix D in einer Weise darstellen, die die Diskussion der OLS-Schätzer (OLS = ordinary least squares) der Regressionskoeffizienten erleichtert: Eine nxn-Matrix D heißt Aggregationsmatrix, wenn sie die folgende Blockdiagonalgestalt besitzt:

$$D = \begin{bmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_m \end{bmatrix} \text{ mit } D_i = 1_{n_i} 1_{n_i}' = \frac{1}{n_i} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{bmatrix} . \tag{2.12}$$

D ist offenbar symmetrisch und idempotent, d.h. $D^2=D$, und somit eine Projektionsmatrix. I_n-D besitzt dann die gleichen Eigenschaften, ist also gleichfalls eine Projektionsmatrix. Insbesondere gilt die Orthogonalitätsbeziehung:

$$(2.13) \qquad (I_n-D)D=D(I_n-D)=0 \ .$$

Offenbar ist nun

$$(2.14) \qquad \bar{X}=DX \ .$$

Somit hat das Modell die zu (2.10) äquivalente Form

$$(2.15) \qquad Y = \begin{bmatrix} 1_n, & X, & DX \end{bmatrix} \begin{bmatrix} \alpha \\ \beta^{(1)} \\ \beta^{(2)} \end{bmatrix} + e \ .$$

Ferner gilt für den Rang von D

$$(2.16) \qquad rg(D) \leq \min(m,K) \ ,$$

woraus folgt, daß für jede nxK-Matrix X gilt:

$$(2.17) \qquad rg(DX) \leq \min(m,K) \ .$$

Bevor wir nun die OLS-Schätzer der Regressionskoeffizienten angeben, sei noch zweierlei angemerkt. Zunächst ist klar, daß der Rang der Designmatrix des Modells keinen vollen Spaltenrang mehr besitzt, sobald $m<K$ ist, also die Anzahl der Gruppen echt kleiner ist als die der unabhängigen Variablen. Die zweite Anmerkung ist prinzipieller Natur. Im vorliegenden Modellansatz behandeln wir $x_{ij}^{(\ell)}$ und $\bar{x}_i^{(\ell)}$ als nichtstochastische Größen; hierbei beschreibt $\bar{x}_i^{(\ell)}$ eine Einflußgröße, die ein eindeutiges Charakteristikum der i-ten Gruppe darstellt. Dies ist nur dann sinnvoll, wenn die i-te Gruppe total erhoben wurde (also aus genau n_i Individuen besteht), da das Charakteristikum sonst nicht aus den vorliegenden Daten bestimmt werden kann. Würde dagegen aus dieser Gruppe etwa nur eine Stichprobe vom Umfang n_i

gezogen, so wäre $\bar{x}_i^{(\ell)}$ bestenfalls noch ein erwartungstreuer Schätzer für das Gruppenmittel; $x_{ij}^{(\ell)}$ und $\bar{x}_i^{(\ell)}$ wären dann als Zufallsgrößen anzusehen. Wir nehmen somit im folgenden stets an, daß die Designmatrix des Modells (2.15) vollen Spaltenrang besitzt, also 2K+1, und daß die Größen $x_{ij}^{(\ell)}$ und $\bar{x}_i^{(\ell)}$ nichtstochastisch sind.

Die Normalgleichungen für die OLS-Schätzung der Regressionskoeffizienten lauten dann in Matrixform

$$
(2.18) \qquad \begin{bmatrix} 1_n'1_n & 1_n'X & 1_n'X \\ X'1_n & X'X & X'DX \\ X'1_n & X'DX & X'DX \end{bmatrix} \begin{bmatrix} \alpha \\ \beta^{(1)} \\ \beta^{(2)} \end{bmatrix} = \begin{bmatrix} 1_n'Y \\ X'Y \\ X'DY \end{bmatrix} ,
$$

die wir mittels einfacher Zeilenoperationen in die folgende Form überführen können:

$$
(2.19) \qquad \begin{bmatrix} n & 1_n'X & 1_n'X \\ 0 & X'(I_n-D)X & 0 \\ 0 & X'(D-M)X & X'(D-M)X \end{bmatrix} \begin{bmatrix} \alpha \\ \beta^{(1)} \\ \beta^{(2)} \end{bmatrix} = \begin{bmatrix} 1_n'Y \\ X'(I_n-D)Y \\ X'(D-M)Y \end{bmatrix} ,
$$

wobei $M = \frac{1}{n} 1_n 1_n'$ die zur Bildung des Gesamtmittelwerts gehörige Aggregationsmatrix ist. Die OLS-Schätzungen der Regressionskoeffizienten der individuellen unabhängigen Variablen lassen sich also aus den Normalgleichungen eines Modells ermitteln, in welches lediglich die Abweichungen der individuellen Variablen von den Gruppenmittelwerten eingehen. Die Summe der geschätzten Koeffizientenvektoren $\beta^{(1)} + \beta^{(2)}$ genügt dagegen den Normalgleichungen eines Modells, in das die Abweichungen der gruppenweisen Mittelwerte vom Gesamtmittelwert eingehen.

Durch die spezielle Konstruktion von $\bar{X}$ aus X wird in der Regel die Multikollinearität der Designmatrix recht hoch sein. Dies hat Konsequenzen bei der numerischen Bestimmung des OLS-Schätzers. Aus diesem Grund wird oft die Matrix X in (2.15) durch die

Matrix $(I_n-D)X$ der Differenzen von den Gruppenmittelwerten ersetzt; wir erhalten dann das Modell

$$(2.20) \qquad Y = \left[1_n,\ (I_n - D)X\ ,\ DX \right] \begin{bmatrix} \tilde{\alpha} \\ \beta^{(1)} \\ \tilde{\beta}^{(2)} \end{bmatrix} + e$$

mit den Normalgleichungen

$$(2.21) \qquad \begin{bmatrix} n & 0 & 1_n' X \\ 0 & X'(I_n-D)X & 0 \\ X'1_n & 0 & X'DX \end{bmatrix} \begin{bmatrix} \tilde{\alpha} \\ \beta^{(1)} \\ \tilde{\beta}^{(2)} \end{bmatrix} = \begin{bmatrix} 1_n'Y \\ X'(I_n-D)Y \\ X'DY \end{bmatrix} ,$$

die sich analog zu den obigen Operationen in die äquivalente Form

$$(2.22) \qquad \begin{bmatrix} n & 0 & 1_n' X \\ 0 & X'(I_n-D)X & 0 \\ 0 & 0 & X'(D-M)X \end{bmatrix} \begin{bmatrix} \tilde{\alpha} \\ \beta^{(1)} \\ \tilde{\beta}^{(2)} \end{bmatrix} = \begin{bmatrix} 1_n'Y \\ X'(I_n-D)Y \\ X'(D-M)Y \end{bmatrix}$$

übertragen lassen. Hier genügen die Regressionskoeffizienten der individuellen und der Kontextvariablen den Normalgleichungen von Modellen, in die jeweils nur individuelle bzw. Kontextvariablen eingehen. Damit stören Kollinearitäten zwischen den Spaltenvektoren von X und $\bar{X}$ nicht mehr bei der numerischen Bestimmung der OLS-Schätzer. Darüber hinaus sind die Schätzer für $\beta^{(1)}$ und $\tilde{\beta}^{(2)}$ auch noch unkorreliert. In der Tat handelt es sich bei (2.20) nur um eine Umparametrisierung von (2.15): Da beide Designmatrizen den gleichen Bildraum besitzen, liefern Identifizierbarkeitsargumente für die beiden Parametersätze

$$(2.23) \qquad \alpha = \tilde{\alpha}\ ;\ \beta^{(1)} = \tilde{\beta}^{(1)};\quad \beta^{(2)} = \tilde{\beta}^{(1)} + \tilde{\beta}^{(2)}\ ;$$

die gleiche Beziehung gilt dann auch für die jeweiligen OLS-Schätzer.

3. Der Boyd-Iversen-Ansatz

Ein zweiter als "klassisch" zu bezeichnender Ansatz, der insbesondere in der Monographie von Boyd und Iversen (1979) untersucht wird, geht von der Annahme aus, daß die Responses zunächst linear durch die Einflußgrößen der Mikroebene gesteuert werden, daß jedoch die Regressionskoeffizienten zufällige Größen sind, die ihrerseits linear durch die Gruppeneffekte reguliert werden. Eine mathematisch exakte Darstellung dieses Ansatzes findet sich in der Arbeit von de Leeuw und Kreft (1986), an die sich die folgenden Ausführungen zunächst recht eng anlehnen und durch deren Erscheinen (zum Leidwesen der Autoren) der Umfang dieses Abschnittes wesentlich verringert werden konnte.

Bezeichnen wir mit y_i den (n_i-dimensionalen) Spaltenvektor der n_i Responses in Gruppe i, so wird als Modell

(3.1) $$y_i = A_i d_i + e_i, \quad E(e_i)=0, \ Kov(e_i) = \sigma_i^2 \ I_{n_i}$$

angenommen, hierbei ist $A_i = [1_{n_i}, X_i]$, wobei X_i definiert ist wie in (2.2); der Vektor $d_i = (d_{0i}, \dots, d_{Ki})'$ der Regressionskonstanten wird dabei nun als zufällig angenommen:

(3.2) $d_1, \dots, d_m$ seien unabhängige ((K+1)-variate) Zufallsvektoren mit $Kov(d_i) =: \Omega$ und $E(d_{\ell 1}, \dots, d_{\ell m})' = B_\ell \beta^{(\ell)}$ für $0 \le \ell \le K$.

Damit ist $\beta^{(\ell)}$ der Vektor derjenigen Regressionskonstanten, der die l-te Komponente der Zufallsgrößen d_i steuert, die Matrix B_ℓ enthält gruppenspezifische Regressoren. Verwenden wir hierfür wieder alle jeweiligen Gruppenmittel der einzelnen unabhängigen Variablen, so ist

(3.3) $$B_\ell = \begin{bmatrix} 1 & \bar{x}_{1.}^{(1)} & \dots & \bar{x}_{1.}^{(K)} \\ \vdots & \vdots & & \vdots \\ 1 & \bar{x}_{m.}^{(1)} & \dots & \bar{x}_{m.}^{(K)} \end{bmatrix} .$$

Schreiben wir nun

(3.4) $$A = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix} \quad \text{und} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

und bilden wir die Vektoren d, e und ß analog zu y, indem wir die Teilvektoren d_i, e_i bzw. $\beta^{(\ell)}$ einfach untereinanderschreiben, so gibt es eine Matrix B derart, daß sich das Modell wie folgt darstellen läßt:

(3.5) $$\begin{aligned} y &= A\,d + e \\ d &= B\,\beta + v\ . \end{aligned}$$

Hierbei sind die Design-Matrizen A und B bekannt (sie enthalten als Komponenten die Einflußgrößen der Mikro- bzw. der Makroebene), ß derjenige unbekannte Parameter, der die Erwartungswerte von d und somit auch von y steuert, und e bzw. v die (als stochastisch unabhängig angenommenen) Fehlerterme, für die sinnvollerweise $E(e)=0$, $E(v)=0$ angenommen wird und deren Kovarianzterme durch die Größen $\sigma_1^2,\ldots,\sigma_m^2$ und Ω spezifiziert sind:

(3.6) $$\text{Kov } e = \text{diag}\,(\sigma_1^2 1'_{n_1},\ldots,\sigma_m^2 1'_{n_m}) \ ; \ \text{Kov } v = \begin{bmatrix} \Omega & & 0 \\ & \ddots & \\ 0 & & \Omega \end{bmatrix}$$

Die genauere Konstruktionsbeschreibung für B findet der interessierte Leser bei de Leeuw und Kreft (1986, Gl.(8)).

Die überlegungen von Boyd und Iversen (1979, Kap.3) laufen nun im wesentlichen auf die Diskussion zweier Schätzverfahren hinaus; wir nehmen hierzu im folgenden stets an, daß die Matrizen A und B vollen Spaltenrang besitzen und somit die Größen A'A, B'B sowie (AB)'(AB) invertierbar sind:

Im "two-step-approach" wird zunächst mittels eines OLS-Schätzers die Realisation $d=(d_1',\ldots,d_m')'$ bestimmt:

$$(3.7) \qquad \hat{d} = (A'A)^{-1}A'y$$

und anschließend $\hat{d}$ als "Beobachtung" in der zweiten Modellgleichung interpretiert und zu einer OLS-Schätzung für β herangezogen:

$$(3.8) \qquad \begin{aligned} \hat{\beta} &= (B'B)^{-1}\, B'\hat{d} \\ &= (B'B)^{-1}\, B'(A'A)^{-1}A'y \ . \end{aligned}$$

Im "single-equation-approach" dagegen werden zunächst die beiden Gleichungen (3.5) zu

$$(3.9) \qquad \begin{aligned} y &= A\ (B\beta + v) + e \\ &= AB\beta + Av + e \\ &= AB\beta + w, \end{aligned}$$

zusammengefaßt, wobei der Fehlerterm w nun die Struktur

$$(3.10) \qquad E(w)=0, \qquad W := \operatorname{Kov} w = \begin{bmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_m \end{bmatrix}$$

$$\text{mit} \quad W_i = \sigma_i^2\, I_{n_i} + A_i \Omega A_i'$$

besitzt. Die OLS-Schätzung von β liefert dann in diesem Fall

$$(3.11) \qquad \mathring{\beta} = (B'A'A\ B)^{-1}\ B'A'y \ .$$

Damit lassen sich die beiden Schätzer kurz darstellen als

$$(3.12) \qquad \hat{\beta} = B^+\ A^+\ y \quad , \quad \mathring{\beta} = (AB)^+\ y \ ,$$

wobei M^+ die Moore-Penrose-Inverse (MPI) der Matrix M bezeichne (vgl. Anhang). Die Eigenschaften der MPI liefern dann sofort, daß die beiden Schätzer genau dann identisch sind, wenn die Matrix $A\ B\ B^+A^+$ symmetrisch ist. Außerdem sieht man sofort, daß beide Schätzverfahren erwartungstreue Schätzer für β liefern.

Es stellt sich natürlich sofort die Frage, ob die Verwendung von $\hat{\beta}$ oder $\mathring{\beta}$ als Schätzer für β wirklich das beste Verfahren darstellt. Wäre die Kovarianz W aus (3.11) bekannt, so würde diese Eigenschaft dem Gauß-Markoff-Aitken-Schätzer

(3.13) $$\beta^* = (B'A'W^{-1}AB)^{-1}B'A'W^{-1}y$$

zukommen. Da W jedoch nicht bekannt ist, liegt es nahe, W durch einen Schätzer $\hat{W}(y)$ zu ersetzen:

(3.14) $$\mathring{\beta}^* = (B'A'\hat{W}(y)^{-1}AB)^{-1}B'A'\hat{W}(y)^{-1}y =: b(y) .$$

De Leeuw und Kreft (1986) sprechen diese Möglichkeit ebenfalls an und konstruieren auch einen erwartungstreuen Schätzer $\hat{W}(y)$ (wir verzichten hier auf die explizite Angabe), sie präzisieren jedoch keine Eigenschaften dieses Schätzers. In der Tat ist die Erwartungstreue von $\mathring{\beta}^*$ nicht selbstverständlich, da $\mathring{\beta}^*$ kein in y linearer Schätzer mehr ist. Tatsächlich gibt es aber in der Literatur Überlegungen, unter welchen Voraussetzungen der Schätzer unverzerrt bleibt: Die einzige weitere Modellannahme, die wir benötigen, ist die Symmetrie der Verteilung von w um 0: w und -w müssen die gleiche Verteilung besitzen, was etwa in dem Fall erfüllt ist, daß wir für die Fehler eine gemeinsame Normalverteilung voraussetzen. Man prüft leicht nach, daß für die in (3.14) definierte Funktion b(y) folgende Bedingungen erfüllt sind:

(3.15) (i) $b(y) = -b(-y)$;
(ii) $b(y + AB\beta) = b(y) + \beta$ für alle β.

Damit sind etwa die Voraussetzungen des Kriteriums von Jöckel (1982:364-365) erfüllt, und es folgt, daß $\mathring{\beta}^*$ um β symmetrisch verteilt ist. Besitzt also $\mathring{\beta}^*$ überhaupt einen Erwartungswert, so ist damit eben wegen dieser Symmetrie auch direkt die Erwartungstreue sichergestellt.

Dieser Beitrag wurde von Michael Wiedenbeck und Günter Rothe verfaßt und ist entstanden aus Vorträgen und Diskussionsbeiträgen während der beiden Workshops über Kontext- und Mehrebenen-

Analyse, die im Mai und September 1986 bei ZUMA abgehalten wurden.

Literatur

Boyd, L.H./G.R. Iversen, 1979: Contextual Analysis: Concepts and Statistical Techniques. Belmont: Wadsworth.

De Leeuw, J./I. Kreft, 1986: Random coefficient models for multilevel analysis. Journal of Educational Statistics 11:57-85.

Erbring, L./A.A. Young, 1980: Individuals and social structure: Contextual effects as endogenous feedback. In: E.F. Borgatta/D.J. Jackson (Hrsg.): Aggregate Data: Analysis and Interpretation. Beverly Hills: Sage.

Jöckel, K.-H., 1982: Iterierte Aitken-Schätzer. Allgemeines Statistisches Archiv 66:361-375.

ANHANG

Eigenschaften der Moore-Penrose-Inversen

(1) Sei M eine beliebige nxk-Matrix. Dann heißt die kxn-Matrix M^+ "Moore-Penrose-Inverse", wenn folgende Eigenschaften erfüllt sind:

(i) MM^+ und M^+M sind symmetrisch

(ii) $MM^+M=M$

(iii) $M^+MM^+ = M^+$.

M^+ ist durch diese Bedingungen eindeutig bestimmt.

(2) Ist in (1) Rang$(M)=k \leq n$, so gilt

$M^+ = (M'M)^{-1}M'$.

In diesem Fall ist dann $M^+M=I$ und $MM^+=P_M$, d.h. die Projektion auf das Bild von M.

(3) Haben A und B vollen Spaltenrang, so gilt $(B^+A^+)\ A\ B = I$, damit erfüllt A^+B^+ alle Bedingungen einer MPI von AB bis auf die erste Symmetrieforderung in (i).