

der Forschung und Publikationsaktivitäten auf die externen Kosten ist dagegen geeignet, die Informationsbasis zu verzerren. Informationsdefizite bei den externen Nutzen des Verkehrs begünstigen allokative Fehlentscheidungen im Hinblick auf die Sicherung des erreichten Wohlstands und im Hinblick auf die Realisierung weiterer Entwicklungsmöglichkeiten. Aus dieser Feststellung ergibt sich der Auftrag an die analytische und an die empirische Wirtschafts- und Verkehrsforschung. Es wäre schon viel gewonnen, wenn die vorhandenen Erkenntnisse zusammengefaßt und ausgewertet würden. Das gilt für die vielen verstreut vorliegenden Untersuchungsergebnisse über produktivitäts- und einkommenssteigernde Impulse von verkehrlichen Investitionen und Leistungsinnovationen. Dann könnte auch ein systematischer und kritischer Methodenvergleich einsetzen. Der Zweck liegt ja nicht darin, ein neues Verwirrspiel mit Zahlen zu beginnen, sondern das zunehmende Wissen so aufzubereiten und zu präsentieren, daß es für eine rationale Argumentation nutzbar wird.

Abstract

The benefit of transport lies in an improvement in the economic conditions brought about by the spatial transfer of people and goods. Since industrialisation, the transport system has been able to achieve a virtually unbroken series of technical and organisational advances. This is a precondition for the sustainable growth of production and living standards. For transport has the quality of a dynamic infrastructure; it enables ever more intensive spatial division of labour in ever bigger markets. Modern transport systems are extraordinarily multi-faceted; they include the services of a number of transport modes whose production, cost and market conditions vary greatly. These systemspecific advantages and disadvantages determine the basic pattern of specialisation and division of labour. For many important service areas there are also substitution possibilities and hence fierce substitution competition. Despite the considerable importance of inland waterways, coastal shipping, pipelines and air transport in certain countries, the relationship between rail and road transport is the crux of the coordination problem of transport within Europe. The system characteristics of the railways offer attractive cost and quality conditions for high-volume transport between nodal points in the economics space; it is a carrier of bulk radial traffics. Road transport on the other hand has undeniable advantages in collection and distribution traffic and more generally in the carriage of relatively small quantities over relatively short distances; in addition, its flexibility makes it possible to meet particular transport needs. These characteristics indicate an essentially complementary relationship of completion and cooperation. The old, simple pattern of division of labour has been completely changed by the economic and social dynamic of the past 30 years however. The sharp increases in productivity and real incomes, together with the shifts in demand structure and production methods, have modified the requirements for the delivery of goods and the mobility of people in such a way that road transport has become the very clearly dominant mode. The most important factors here were the technical and organisational possibilities for flexible adjustment of the service provided to the new demands of logistic systems planning. The extremely rapid expansion of road passenger and freight transport has however led to increasing disamenity and environmental pollution and to congestion caused by bottlenecks in the road network. The appearance of high external costs has given to criticism of this development. Acceptance of the costs of environmental pollution and road accidents would have weakened the competitive position of the railways. The dominant position of road transport thus would not reflect the "true" benefit contribution of the transport modes. Explicit and fairly one-sided concentration on the external costs has led to the question of the benefits of transport, and especially of road transport, being seen as a problem of market regulation. Although there is a broad consensus on the importance of road transport in the economic expansion of recent decades, there is argument about the interpretation and imputation of these benefits. The essential question is whether the benefits of transport services are fully and completely rewarded through the market, or whether there are external benefits as well as internal. In this paper we try to demonstrate the appearance of external benefits of transport and emphasize their relevance for the allocation of the factors of production. The provision of transport services creates development promoting system benefits, which, because of their nature and the dispersion of the effects can never be fully imputed to the provider and paid for. A similar state of affairs is found with the implementation of important discoveries or the diffusion of the benefits of investment in education. The appearance of external costs is therefore not a sufficient criterion for saying that there is corresponding need for internalisation. The experience and expectation of external benefits in practice leads rather to external costs being accepted to a certain extent as a kind of price. Under abstract model assumptions a tendency towards equilibrium between external benefits and costs can be deduced. This approach makes it possible to reformulate the question of the optimal extent of internalisation of external costs.

Verkehrsdatenmix und Zeitcluster - ein Ansatz zur multivariaten Verkehrsanalyse

VON KLAUS-JÜRGEN RICHTER, DRESDEN

v. st. a
v. fe. a
S

Inhalt

1. Einführung
2. Zeiteinheiten als „Objekte“
3. Verkehrsdatenmix
4. Dynamische Distanzmessung
5. Schlußbemerkungen

1. Einführung

Im folgenden Beitrag wird versucht, die Methodologie und die Verfahren der multivariaten Statistik, genauer: einiger Teile der multivariaten Statistik, auf die gemeinsame Analyse mehrerer Zeitreihen anzuwenden. Dieser Sachverhalt dürfte in der Verkehrsanalyse oftmals vorliegen, in der es in aller Regel nicht genügt, eine einzelne Zeitreihe oder mehrere Zeitreihen einzeln zu analysieren. Die verwendeten statistischen Urdaten betreffen das Gebiet der neuen Bundesländer, also die ehemalige DDR.[1] Sie dienen vor allem der datenmäßigen Unterstützung der vorgeschlagenen Vorgehensweise; bei einer praktischen Verkehrsanalyse wären zweifellos mehr und wahrscheinlich auch noch andere Daten einzubeziehen.

2. Zeiteinheiten als „Objekte“

In der multivariaten Statistik, insbesondere bei der Aufstellung der multivariaten Datei, spielen die Begriffe des Objekts und des Merkmals eine zentrale Rolle. [2, 3, 4, 5, 6] Dabei ist ein Objekt die Elementareinheit einer definierten statistischen Masse oder Gesamtheit, die somit aus n Objekten besteht. An jedem Objekt werden die gleichen p Eigenschaften, die als Merkmale bezeichnet werden, untersucht.

Die Ausgangsdaten einer multivariaten statistischen Analyse sind somit gekennzeichnet durch

- n Objekte O_i , $i = 1(1)n$,
- p Merkmale M_j , $j = 1(1)p$,
- $n \cdot p$ Meßwerte x_{ij} , $i = 1(1)n$,
 $j = 1(1)p$.

Anschrift des Verfassers:

Prof. Dr. Klaus-Jürgen Richter
Leiter des Instituts für Wirtschaftsinformatik
an der Hochschule für Verkehrswesen „Friedrich List“
Friedrich-List-Platz 1
O - 8010 Dresden

Die Merkmalsrealisation des Merkmals M_j , die am Objekt O_i festgestellt wurde, heißt x_{ij} . Alle Merkmalsrealisationen zusammen werden zweckmäßigerweise in der Datenmatrix \underline{X} zusammengestellt:

$$\underline{X} = (x_{ij})_{np} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

Ersetzt man die Objekte durch aufeinanderfolgende Zeiteinheiten, die in aller Regel äquidistant sind, so ändert sich an der vorgestellten formalen Grundstruktur nichts. Die Matrixspalten bleiben nach den Merkmalen geordnet, während die Matrixzeilen nicht mehr durch die Untersuchungsobjekte bestimmt sind, sondern durch die Zeiteinheiten. Zeiteinheiten können dabei sowohl Zeitpunkte als auch Zeiträume sein.

Inhaltlich tritt allerdings eine gewisse Umkehrung des ursprünglichen Ansatzes der multivariaten Analyse ein: Während es zunächst darum ging, eine bestimmte Gruppe von Objekten nach einer Menge gemeinsamer Merkmale zu kennzeichnen, wird nunmehr die gemeinsame Veränderung einer Merkmalsgruppe in der Zeit betrachtet. Das verlangt die präzise Abgrenzung des einen Objektes, auf das sich alle Merkmale beziehen. Im Falle der Verkehrsanalyse kann es sich dabei um eine regionale oder allgemein territoriale Einheit, aber auch um ein Verkehrsunternehmen handeln.

In der folgenden Tabelle 1 sind für das einleitend genannte Untersuchungsgebiet und für die Jahre 1980 bis 1989 folgende Merkmale zusammengestellt worden:

- X_1 Wohnbevölkerung in Mill. Personen
- X_2 Produziertes Nationaleinkommen in Mrd. Mark
- X_3 Gütertransportmenge in Mrd. Tonnen
- X_4 Gütertransportleistung in Mrd. Tonnenkilometern
- X_5 Beförderte Personen in Mrd. Personen
- X_6 Personenbeförderungsleistung in Mrd. Personenkilometern
- X_7 Bestand an zugelassenen Personenkraftwagen in Mill. Pkw

Einige der in Tabelle 1 enthaltenen Zeitreihen sind praktisch stationär, einige nur schwach nichtstationär. Durch die Datentransformation wird es jedoch möglich sein, die Unterschiede zwischen den einzelnen Merkmalen deutlicher hervorzuheben. Wesentliche Bezugsgrößen jeder Datentransformation sind der Durchschnitt und die Standardabweichung für jedes Merkmal, die als

Tabelle 1: Ausgangsdaten [1, S. 1, 13, 248, 249, 252] (gegenüber der Quelle gerundet)

Jahr	i	X_1	X_2	X_3	X_4	X_5	X_6	X_7 *)
1980	1	16,7	193,6	1,1	155,3	4,1	53,9	2,7
1981	2	16,7	203,0	1,1	154,0	4,1	53,2	2,8
1982	3	16,7	208,2	1,0	133,0	4,1	52,7	2,9
1983	4	16,7	217,8	1,0	142,4	4,1	53,1	3,0
1984	5	16,7	230,0	1,0	144,1	4,2	54,1	3,2
1985	6	16,7	242,0	1,0	144,8	4,2	54,5	3,3
1986	7	16,6	252,2	1,0	157,0	4,1	54,9	3,5
1987	8	16,7	260,6	0,9	173,6	4,2	55,7	3,6
1988	9	16,7	268,1	1,0	171,8	4,2	56,8	3,7
1989	10	16,4	273,7	1,0	174,9	4,1	57,6	3,9

*) darunter mehr als 90% private Personenkraftwagen.

$$\bar{x}_j = \frac{1}{n} \sum_i x_{ij} \quad \forall j \quad (2)$$

und

$$s_j = \sqrt{\frac{1}{n} \sum_i (x_{ij} - \bar{x}_j)^2} \quad \forall j \quad (3)$$

berechnet und in Tabelle 2 zusammengestellt werden.

Tabelle 2: Durchschnitt und Standardabweichung für jedes Merkmal nach Tabelle 1

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
\bar{x}_j	16,66	234,9	1,01	155,1	4,14	54,7	3,26
s_j	0,09	27,10	0,05	13,80	0,05	1,54	0,39

Die Matrix der Ursprungsdaten \underline{X} nach (1) kann entsprechend den Regeln der multivariaten Analyse in Spaltenvektoren und in Zeilenvektoren zerlegt werden.

Die Zerlegung in Spaltenvektoren ergibt

$$\underline{X} = (\underline{x}_j)_p \tag{4}$$

Jeder Spaltenvektor

$$\underline{x}_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ \vdots \\ x_{nj} \end{bmatrix} \quad \forall j \tag{5}$$

liefert die zeitlich geordneten Daten für ein Merkmal und stellt somit eine Zeitreihe im üblichen Sinne dar. Die Elemente eines solchen Vektors lassen sich nach den statistischen Prinzipien der Zeitreihenanalyse bearbeiten.

Jeder Zeilenvektor

$$\underline{x}(i) = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}] \quad \forall i \tag{6}$$

enthält die Werte, die für alle Merkmale für eine Zeiteinheit i gefunden worden sind. Er liefert auf der Basis der gewählten Merkmale eine komplexe Charakteristik der Zeiteinheit i . Alle Zeilenvektoren bilden entsprechend

$$\underline{X} = \begin{bmatrix} \underline{x}(1) \\ \underline{x}(2) \\ \underline{x}(3) \\ \vdots \\ \underline{x}(n) \end{bmatrix} \tag{7}$$

wieder die Datenmatrix \underline{X} .

Zur Bearbeitung einer multivariablen Datei werden verschiedene Arten der Datentransformation herangezogen. Das gilt auch für die hier erörterte dynamische multivariate Datei.

Vor allem handelt es sich um

- die Mittelwertkorrektur,
- die Standardisierung und
- die Linearkombination.

Auf der Grundlage der Standardisierung können Kreuzproduktmatrizen berechnet werden, von denen die Korrelationsmatrix besondere Bedeutung besitzt. Die Linearkombination wird benötigt, um den Verkehrsdatenmix zu bestimmen.

Die Mittelwertkorrektur besteht darin, daß von allen Elementen eines Spaltenvektors der Datenmatrix \underline{X} der zugehörige Durchschnitt subtrahiert wird. Aus den Elementen der Matrix \underline{X} , x_{ij} , entstehen die mittelwertkorrigierten Elemente

$$x_{ij.d} = x_{ij} - \bar{x}_j \quad \forall i,j \tag{8}$$

die ebenfalls in einer entsprechenden Matrix \underline{X}_d zusammengestellt werden können. Die Elemente jedes Spaltenvektors dieser Matrix besitzen den Durchschnitt Null.

Durch Division der Werte nach (8) mit der Standardabweichung s_j , jeweils spaltenweise vorgenommen, entstehen die standardisierten Daten $x_{ij.s}$ als

$$x_{ij.s} = \frac{x_{ij.d}}{s_j} \quad \forall i,j \tag{9}$$

Sie ergeben gemeinsam die Matrix \underline{X}_s . Die Elemente jedes Spaltenvektors dieser Matrix besitzen den Durchschnitt Null und die Standardabweichung Eins. Nach der Anzahl der Zeilen und der Spalten stimmen die Matrizen \underline{X} , \underline{X}_d und \underline{X}_s überein. Tabelle 3 enthält die standardisierten Werte gemäß (9).

Tabelle 3: Standardisierte Merkmalswerte, berechnet auf der Grundlage von Tabelle 1

i	$X_{1.s}$	$X_{2.s}$	$X_{3.s}$	$X_{4.s}$	$X_{5.s}$	$X_{6.s}$	$X_{7.s}$
1	0,44	-1,52	1,80	0,01	-0,80	-0,48	-1,43
2	0,44	-1,18	1,80	-0,08	-0,80	-0,94	-1,18
3	0,44	-0,98	-0,20	-1,60	-0,80	-1,27	-0,92
4	0,44	-0,63	-0,20	-0,92	-0,80	-1,01	-0,67
5	0,44	-0,18	-0,20	-0,79	1,20	-0,36	-0,15
6	0,44	0,26	-0,20	-0,74	1,20	-0,10	0,10
7	-0,66	0,64	-0,20	0,14	-0,80	0,16	0,61
8	0,44	0,95	-2,20	1,34	1,20	0,68	0,87
9	0,44	1,22	-0,20	1,21	1,20	1,40	1,13
10	-2,88	1,43	-0,20	1,43	-0,80	1,92	1,64

Tabelle 3 korrespondiert natürlich mit Tabelle 1. Durch die vorgenommene Standardisierung der Daten enthält sie insgesamt sieben Zeitreihen dimensionsloser Werte, die untereinander leicht vergleichbar sind, sofern man dazu die Abweichung vom jeweiligen Durchschnitt nach Richtung und Betrag, gemessen in Einheiten der jeweiligen Standardabweichung, heranzieht. Die relativ sprunghafte Veränderung der Werte in einigen Spalten ist das Ergebnis der starken Rundungen, denen die Ursprungswerte unterworfen wurden. Sie ist bei weniger starker Rundung vermeidbar. Der Verlauf der Werte nach Tabelle 3 ist in Abbildung 1 dargestellt.

Es wurde bereits darauf hingewiesen, daß von den verschiedenen Arten der Kreuzproduktmatrizen die Korrelationsmatrix besondere Bedeutung besitzt. Sie enthält die Korrelationskoeffizienten aller Merkmale untereinander, die in der jeweiligen multivariaten Analyse auftreten. Das gilt auch für die dynamische Analyse. Somit gilt für diese Matrix allgemein

$$\underline{R} = (\tau_{jk})_{pp} \quad (10)$$

Die Matrix enthält in der Hauptdiagonalen die Autokorrelationskoeffizienten. Außerdem besitzt sie die Symmetrieeigenschaft, indem

$$\tau_{jk} = \tau_{kj} \quad (11)$$

für alle Werte von j und k gilt. Bei insgesamt p Merkmalen ergibt sich somit allgemein:

$$\underline{R} = \begin{bmatrix} 1 & \tau_{12} & \tau_{13} & \dots & \tau_{1p} \\ & 1 & \tau_{23} & \dots & \tau_{2p} \\ & & 1 & \dots & \tau_{3p} \\ & & & \dots & \dots \\ & & & & \dots \\ & & & & & 1 \end{bmatrix} \quad (12)$$

Für die sieben Merkmale der vorliegenden Untersuchung wurde \underline{R} berechnet. Die damit bestimmte Korrelationsmatrix ist der besseren Übersichtlichkeit wegen als Tabelle 4 angegeben.

Auf eine nähere inhaltliche Analyse der gefundenen Werte wird an dieser Stelle verzichtet. Es wird jedoch bereits auf den ersten Blick deutlich, daß die einbezogenen Merkmale nach Stärke und Richtung sehr unterschiedlich miteinander korrelieren.

Die durch die Korrelationsmatrix \underline{R} vermittelte Aussage bestätigt sich durch den Verlauf der standardisierten Werte aller sieben Merkmale bzw. Variablen in Abbildung 1. Um die durch die Standardisierung gewonnene Information beurteilen zu können, stelle man sich vor, die ursprünglichen Merkmalswerte nach Tabelle 1 seien über der Zeitachse aufgetragen worden.

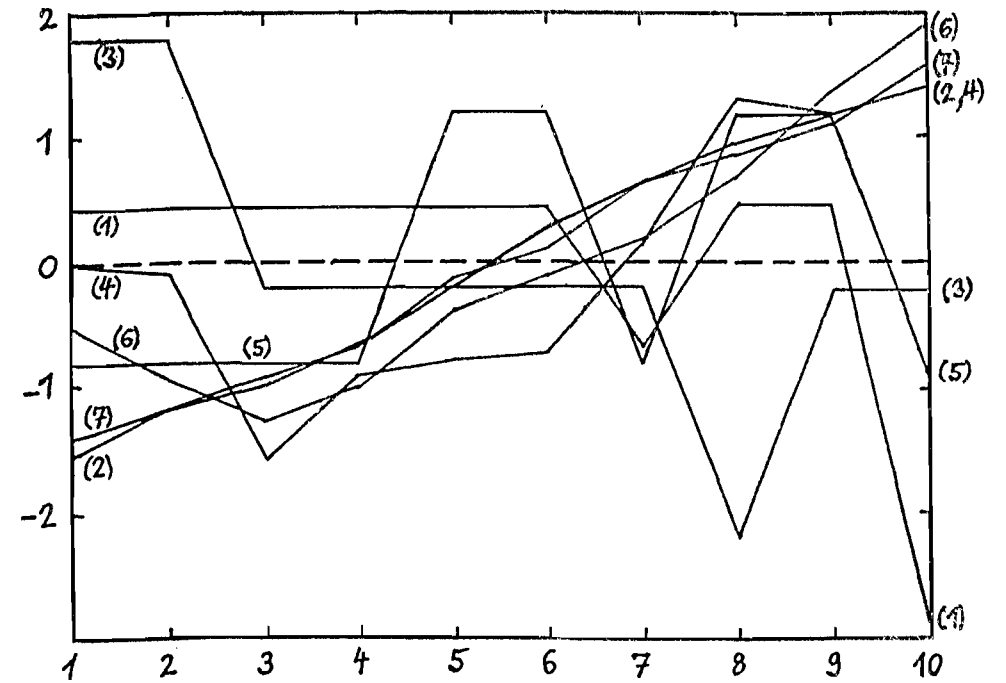


Abbildung 1: Verlauf der Werte nach Tabelle 3

Herausgekommen wären zwei Arten von Verlaufsformen, nämlich drei quasi stationäre Punktfolgen und vier mit einem mehr oder weniger starken Anstieg. Die Standardisierung erweist sich als eine Möglichkeit, den Charakter von Entwicklungsprozessen zu verdeutlichen. Daneben bildet sie einen Zugang zur Bildung des Verkehrsdatenmix in Form einer Linearkombination der ursprünglichen Merkmale.

Tabelle 4 - Korrelationsmatrix

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
X ₁	1	-0,5447	0,0884	-0,4902	0,3536	-0,6551	-0,6116
X ₂		1	-0,7302	0,7061	0,4492	0,8961	0,9884
X ₃			1	-0,2820	-0,5600	-0,4200	-0,6960
X ₄				1	0,2040	0,8762	0,7176
X ₅					1	0,3240	0,3900
X ₆						1	0,9052
X ₇							1

Wie man sieht, bestehen hinsichtlich der Verfahrensweisen und der Interpretation der Ergebnisse keine grundsätzlichen Probleme, wenn die Objekte der multivariaten Statistik durch Zeiteinheiten ersetzt werden. Es ist allerdings zu beachten, daß die angewandten statistischen Maßzahlen wie z. B. Durchschnitt und Standardabweichung dadurch, daß sie auf Zeitreihen bezogen werden, eine Spezifizierung erhalten. Jedoch werden damit die Möglichkeiten der dynamischen multivariaten Analyse nicht eingeschränkt.

3. Verkehrsdatenmix

Eine Form der Datentransformation besteht in der Verknüpfung mehrerer Merkmale bzw. Variablen miteinander zu einer Linearkombination. Bei Vorhandensein standardisierter Variablen ist eine Linearkombination durch

$$y_{i.s} = \sum_j w_j x_{ij.s} \quad \forall i \quad (13)$$

gegeben. Darin bedeuten

$y_{i.s}$ die zusammengesetzte (kombinierte) Variable für das Objekt bzw. die Zeiteinheit i , gebildet aus den standardisierten Werten der einzelnen Variablen,

w_j das Gewicht (die Bedeutung) der Variablen X_j in der Linearkombination und

$x_{ij.s}$ der standardisierte Wert des Merkmals j am Objekt (an der Zeiteinheit) i .

Die Gewichtszahlen werden im allgemeinen normiert:

$$0 \leq w_j \leq 1 \quad \forall j \quad (14)$$

$$\sum_j w_j = 1 \quad (15)$$

Die Festlegung der Gewichtszahlen bereitet oft Schwierigkeiten. Deshalb wird im allgemeinen gefordert, in die multivariate Analyse Variablen bzw. Merkmale mit annähernd gleicher Bedeutung einzubeziehen. Das läuft darauf hinaus, daß bei p Merkmalen jedem Merkmal der Bedeutungswert

$$w_j = 1/p \quad \forall j \quad (16)$$

zukommt. In diesem Falle kann auf Einhaltung von Bedingung (15) verzichtet und einfach

$$w_j = 1 \quad \forall j \quad (17)$$

gesetzt werden. Das wirkt sich ausschließlich auf den Zahlenwert von $y_{i.s}$ aus. Die Tatsache, daß die Linearkombination eine synthetische Größe ist, die in der Regel keine sinnvolle semantische Belegung erfahren kann, wird dadurch nicht berührt.

In Formel (13) wurden die standardisierten Variablenwerte in die Linearkombination einbezogen. Das ist nicht zwingend; möglich ist es ebenfalls, ursprüngliche oder mittelwertkorrigierte Variablenwerte zu benutzen.

Kombiniert man gem. (13) Verkehrsdaten und/oder verkehrsrelevante Daten, so entsteht ein Verkehrsdatenmix, wobei alle allgemeinen Aussagen zur Linearkombination gelten. Tabelle 5 enthält dafür ein Beispiel.

Tabelle 5: Verkehrsdatenmix (gebildet aus den standardisierten Werten der Tabelle 3 mit den einheitlichen Gewichten $w_j = 1$)

Jahr	i	Verkehrsdatenmix aus	
		X_1 bis X_7	X_2 bis X_7
1980	1	-1,98	-2,42
1981	2	-1,94	-2,38
1982	3	-5,33	-5,77
1983	4	-3,79	-4,23
1984	5	-0,04	-0,48
1985	6	0,96	0,52
1986	7	-0,11	0,55
1987	8	3,28	2,84
1988	9	6,40	5,96
1989	10	2,54	5,42

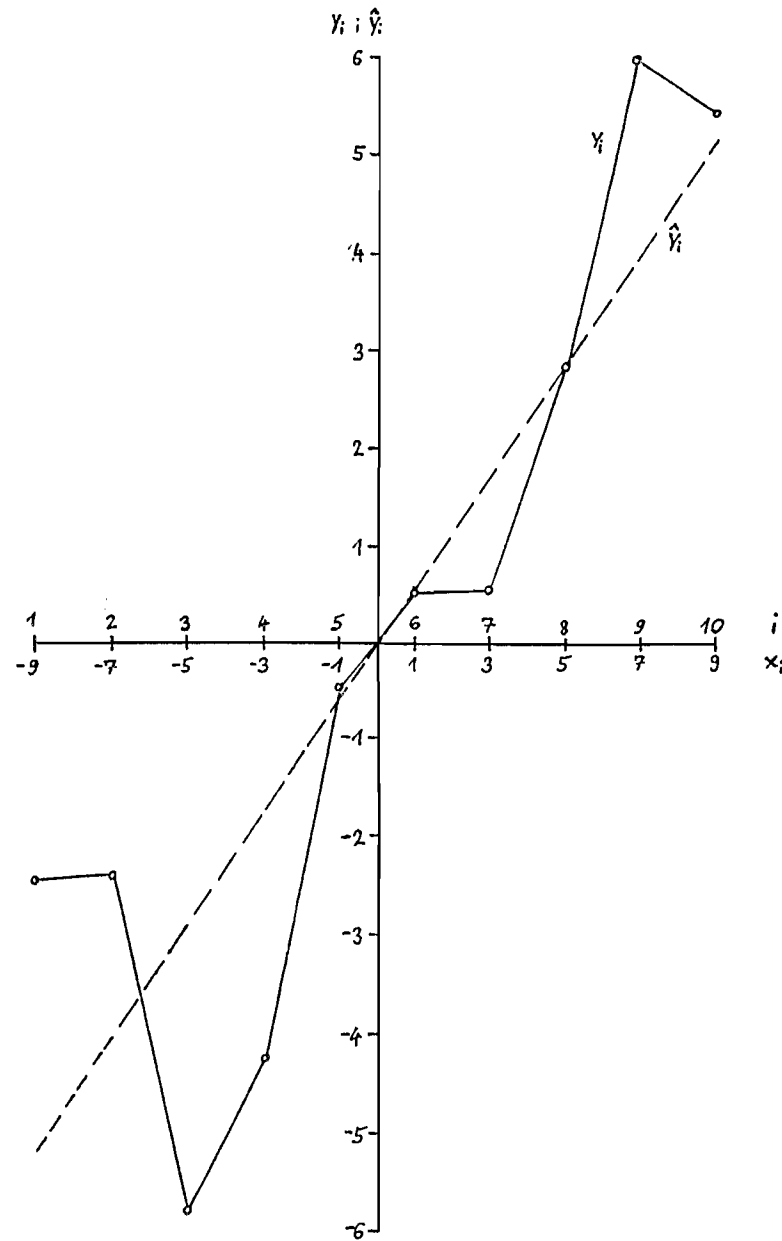
Bei zweifache Berechnung wurde vorgenommen, um den störenden Einfluß der Variablen $X_{1.s}$ im Verkehrsdatenmix auszuschalten. Infolge der vorgenommenen Standardisierung bei vergleichsweise geringer Streuung, die wiederum auf die überwiegende Gleichheit der Ausgangswerte zurückzuführen ist, führen geringe Abweichungen in den Ausgangswerten zu extremen Abweichungen in den standardisierten Daten, die im Verkehrsdatenmix möglichst vermieden werden sollen.

In Abbildung 2 sind die Datenmixwerte nach Tabelle 5, letzte Spalte, und die zugehörigen Werte der linearen Trendfunktion eingetragen worden. Die Trendfunktion lautet:

$$\hat{y}_i = 0,001 + 0,568 x_i \quad (18)$$

Die x_i -Wertefolge wurde aus der Zeigerfolge der i -Werte durch Transformation hergeleitet.

Der Verkehrsdatenmix stellt eine synthetische Variable dar, die auch nur als solche interpretiert werden kann. Somit muß der Nutzer einer solchen Variablen wissen, aus welchen ursprünglichen Variablen sich y_i zusammensetzt (der zweite Index s nach Formel (13) wurde weggelassen, weil davon ausgegangen wird, daß in y_i standardisierte Variablen zusammengefaßt werden). Wenn vorausgesetzt wird, daß in aller Regel ursprüngliche Merkmale bzw. Variablen von annähernd gleicher Bedeutung vereinigt werden, so läßt sich eine solche synthetische Variable für verschiedene Bewertungen und Vergleiche, dabei auch für den Zeitvergleich und die Darstellung von Entwicklungen verwenden.

Abbildung 2: Datenmixwerte (ohne X_i) und linearer Trend

Daneben ist es natürlich möglich und für untersetzende Aussagen oft auch nützlich, die einzelnen Zeitreihen oder einige von ihnen, die in y_i eingegangen sind, getrennt zu analysieren.

Es bedeutet eine mehr verkehrswirtschaftliche als statistische Frage, welche Merkmale zum Verkehrsdatenmix zu vereinigen sind. Die Beantwortung dieser Frage hängt nicht zuletzt von empirischen Untersuchungen ab, bei denen dieser Mix im zeitlichen und eventuell auch im zeitlich-regionalen Vergleich eingesetzt wurde.

4. Dynamische Distanzmessung

In der Clusteranalyse [5, 6], einer der leistungsfähigen Methoden der multivariaten statistischen Analyse, werden die Abstände oder Distanzen zwischen allen in die Analyse einbezogenen Objekten auf der Grundlage aller Merkmale (Variablen), die erfasst worden sind, bestimmt. Diese Distanzen werden benutzt, um die Objekte zu Gruppen, sog. Clustern zusammenzufassen. Cluster sollen dergestalt gebildet sein, daß die Objekte innerhalb eines Clusters einander sehr ähnlich sind, also zwischen ihnen nur geringe Distanzen bestehen, während zwischen den Clustern deutliche bzw. signifikante Unterschiede erkennbar sein sollen. Die Clusterbildung dient somit der sachlich begründeten Strukturierung einer Objektmenge.

Die Clusterung erfolgt in zwei Schritten, nämlich

- a) in der Bestimmung aller Distanzen zwischen allen Objekten und
- b) in der Bildung der Cluster selbst.

Da für die dynamische Distanzmessung der Schritt b) geringere Bedeutung besitzt als bei der Clusterung von Objekten, konzentrieren wir uns auf Schritt a). In der relativ kurzen Entwicklungsgeschichte der Clusteranalyse sind verschiedene Distanzmaße vorgeschlagen worden, von denen beispielsweise

- die Euklidische Distanz,
- die City-Block-Distanz und
- die Minkowski-Metrik

genannt werden. [5, 6] Das erste und das zweite Distanzmaß sind Sonderfälle des dritten. Nachfolgend wird die Euklidische Distanz benutzt. Nach ihr ergibt sich der Abstand zwischen den Objekten O_i und O_k als

$$d_{i.k.s} = \sqrt{\sum_j (x_{i.j.s} - x_{k.j.s})^2} \quad (19)$$

In die Distanzmessung werden alle p Merkmale oder Variablen einbezogen. Es werden standardisierte Werte benutzt. Es wird aber der Einfachheit halber anschließend nur $d_{i.k}$ anstatt $d_{i.k.s}$ geschrieben. Sind p Merkmale definiert, so können selbstverständlich auch Distanzen auf der Grundlage der einzelnen Merkmale oder von Merkmalsgruppen gebildet werden. Alle nach (19) bestimmten Distanzen werden in der Distanzmatrix \underline{D} zusammengestellt:

Die aufgestellten Distanzmatrizen, insbesondere D^x , erlauben eine dynamische Clusterung im doppelten Sinne:

- a) Es ist möglich, nach der herkömmlichen Vorgehensweise der Clusteranalyse Gruppen bzw. Cluster „ähnlicher“ Zeiteinheiten zu bilden, und zwar unabhängig von der zeitlichen Reihenfolge bzw. der zeitlichen Nähe der Zeiteinheiten. Das ermöglicht die Aufdeckung von Ähnlichkeiten zwischen Zeiteinheiten auf multivariater Basis. Ein entsprechendes Drendrogramm kann aufgestellt werden. [5, 6]
- b) Innerhalb einer Zeitreihe können unterschiedliche Veränderungstempi auftreten, die sich durch entsprechende Abstände zwischen benachbarten Zeiteinheiten ausdrücken. Eng beieinander liegende Zeiteinheiten, wie beispielsweise in Tabelle 7 die Einheiten 1 und 2, 3 und 4 sowie 5 und 6 bilden Minicluster in Zeitfolge.

Beide Wege lassen sich bis zur Darstellung von Periodizitäten im allgemeinen und im verkehrswirtschaftlichen Ablauf führen.

Bei der Distanzbemessung nach Formel (19) werden alle Merkmale mit gleichem Gewicht eingesetzt. Das entspricht der Art und Weise, wie die Werte des Verkehrsdatenmix bestimmt worden sind. Dann liegt der Gedanke nahe, diese Mixwerte unmittelbar heranzuziehen, um die Differenzen zwischen ihnen als Maß für die Distanzen zwischen den Zeiteinheiten zu benutzen. Man erhält Tabelle 8.

Tabelle 8: Distanzmatrix auf der Grundlage der Werte des Verkehrsdatenmix (Tabelle 5, letzte Spalte)

	1	2	3	4	5	6	7	8	9	10
1	0	-0,04	3,35	1,81	-1,94	-2,94	-2,97	-5,26	- 8,38	- 7,84
2		0	3,39	1,85	-1,90	-2,90	-2,93	-5,22	- 8,34	- 7,80
3			0	-1,54	-5,29	-6,29	-6,32	-8,61	-11,73	-11,19
4				0	-3,75	-4,75	-4,78	-7,07	-10,19	- 9,65
5					0	-1,00	-1,03	-3,32	- 6,44	- 5,90
6						0	-0,03	-2,32	- 5,44	- 4,90
7							0	-2,29	- 5,41	- 4,87
8								0	- 3,12	- 2,58
9									0	0,54
10										0

Die auf die gleiche Weise bestimmten Distanzen zwischen den unmittelbar benachbarten Zeiteinheiten stehen in Tabelle 9.

Die in den Tabellen 8 und 9 enthaltenen Distanzen sind als

$$d_{ik}^x = y_i - y_k \quad \text{mit } i \leq k \quad (27)$$

gebildet worden, worin y_i und y_k Datenmixwerte sind. Es handelt sich um lineare Größen, die die Distanz zwischen den Zeiteinheiten i und k nach Betrag und Richtung (Vorzeichen) angeben. Infolge dieser Eigenheiten gelten folgende Beziehungen:

Tabelle 9: Reduzierte Distanzmatrix zu Tabelle 8

	1	2	3	4	5	6	7	8	9	10
1	0	-0,04								
2		0	3,39							
3			0	-1,54						
4				0	-3,75					
5					0	-1,00				
6						0	-0,03			
7							0	-2,29		
8								0	-3,12	
9									0	0,54
10										0

$$d_{ki}^x = (-1) d_{ik}^x \quad (28)$$

$$d_{ii}^x = 0 \quad (29)$$

$$d_{ik}^x = d_{i1}^x + d_{1k}^x \quad (30)$$

Ein Vergleich mit den Formeln (23) ff. zeigt die Unterschiede zur Euklidischen Distanz. Wie auf der Grundlage der Tabellen 6 und 7 können auch auf derjenigen der Tabellen 8 und 9 Clusterungen vorgenommen werden.

Wie in [7] ausführlich gezeigt wird, lassen sich auf der Grundlage der multivariaten Zeitreihenanalysen umfassende Modelle aufbauen, die sowohl für die Verkehrsanalyse als auch für die Verkehrsprognose geeignet sein können. Angesichts der methodisch oft schwach abgesicherten Verkehrsprognosen verdient dieser Umstand besondere Erwähnung.

5. Schlußbemerkungen

Es wurde der Versuch unternommen, in Anbetracht der Komplexität verkehrswirtschaftlicher Analysen und der dazu erforderlichen statistischen Absicherung die Vorgehensweise der multivariaten Statistik auf Zeitreihenprobleme zu übertragen. Dabei hat sich gezeigt, daß den Objekten der multivariaten Analyse im Bereich der dynamischen Analyse die Zeiteinheiten entsprechen. Mittels Linearkombination läßt sich ein Verkehrsdatenmix, allgemein ein Datenmix beliebiger Zusammensetzung, aufbauen, wobei die Standardisierung der Ursprungsdaten vorausgesetzt wird. Semantik und Aussagefähigkeit eines Verkehrsdatenmix hängen primär von der richtigen Auswahl der konstituierenden Ausgangsdaten ab. Der in der Clusteranalyse fundamentale Begriff der Distanz ist auch im Bereich der Zeitreihenanalyse verwendbar, und zwar sowohl in der ursprünglichen Form als auch als lineare

Distanz auf der Grundlage der Mixwerte. Die so gebildeten Distanzen lassen sich zur Clustierung heranziehen.

Bei der Interpretation der Ergebnisse, die hier am Beispiel weniger Daten nur angedeutet wurde, ist die unterschiedliche Ausgangssituation der Zeitreihenanalyse gegenüber der herkömmlichen Statistik zu beachten, um Fehldeutungen auszuschließen.

Literatur

- [1] Statistisches Jahrbuch der Deutschen Demokratischen Republik 1990. Berlin 1990.
- [2] *Backhaus, K.* u. a.: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 5., rev. Auflage. Berlin, Heidelberg, New York 1989.
- [3] *Fahrmeir, L.; A. Hamerle* (Hrsg.): Multivariate statistische Verfahren. Berlin, New York 1984.
- [4] *Hartung, J.; B. Elpelt*: Multivariate Statistik. Lehr- und Handbuch der angewandten Statistik. 3., durchgeseh. Aufl. München, Wien 1989.
- [5] *Steinhausen, D.; K. Lange*: Clusteranalyse. Einführung in Methoden und Verfahren der numerischen Klassifikation. Berlin, New York 1977.
- [6] *Eckes, T.; H. Roßbach*: Clusteranalysen. Stuttgart 1980.
- [7] *Lütkepohl, H.*: Introduction to Multiple Time Series Analysis. Berlin, Heidelberg, New York 1991.

Abstract

Multivariate statistics include a few of methods to analyse some variables in a complex. This situation for example we also can find in studies of transport economics. Some approaches to combine methods of multivariate statistics with the idea of time series analysis are proposed in the paper. So it is possible to use multivariate objects like time units. Different single variables can be linked together to a mix of economic transport datas. Final the time units are clustered like general objects. A special type clustering is based on the data mixes. The used statistical values describe some aspects of the situation in transport of the so called new countries of GFR from 1980 to 1989.

Aufbau, Methoden und Ziele der Fahrleistungserhebung 1990

VON GÜNTER ZIMMERMANN, BERGISCH GLADBACH

v. l. e. a
v. st. a
s

1. Ausgangslage

Themen und Diskussionen verkehrsstatistischer Seminare in den 80er Jahren beschäftigten sich vor allem mit dem Mangel an verlässlichen Informationen über Grundphänomene der verschiedenen Verkehrsarten. Vertreter der Fachöffentlichkeit forderten übereinstimmend eine intensive Weiterentwicklung der Erhebungs- und Analysemethoden in der Verkehrsstatistik. Gültigkeit und Zuverlässigkeit verkehrswissenschaftlicher Theorien und daraus abgeleiteter system- sowie prozessorientierter Modelle sollten anhand empirisch gewonnener Daten überprüft werden. Spezielles Gewicht erhielt dabei die Forderung nach einer Neubestimmung der fahrzeugbezogenen Fahrleistungswerte.

Die Datenlage war dadurch gekennzeichnet, daß die letzten amtlichen Fahrleistungserhebungen in den Jahren 1959 und 1966 durchgeführt worden waren und die für das Jahr 1974 geplante Erhebung aufgrund einer parlamentarisch festgelegten Straffung der Bundesstatistik gestrichen wurde.

Zwischenzeitlich vorgenommene Schätzungen privater und öffentlicher Institutionen waren aus Gründen unterschiedlicher Definitionen und Methoden nicht vergleichbar. Deshalb beauftragte die Bundesanstalt für Straßenwesen (BASt) eine Projektgruppe mit der Erstellung eines Berichtes über Inhalt, Methoden und Instrumente einer umfassenden Fahrleistungserhebung und über Möglichkeiten einer effizienten Realisierung. Der Ende 1978 vorgelegte Bericht bildete die Grundlage für den „Entwurf eines Gesetzes über eine Statistik der Fahr- und Beförderungsleistungen von Kraftfahrzeugen“, der im Jahr 1982 vom Bundesminister für Verkehr in Zusammenarbeit mit anderen Datennutzern erarbeitet wurde, aber nicht in den parlamentarischen Gesetzgebungsprozeß gelangte.

Auch die bereits erwähnten dringlichen Forderungen der Fachleute konnten die Realisierungschancen für ein „Bundesfahrleistungsgesetz“ nicht erhöhen. In dieser Situation stimmte der Bundesminister für Verkehr zu, in einem Forschungsprogramm der BASt eine bundesweite Fahrleistungserhebung für das Jahr 1990 auszuweisen.

2. Voraussetzungen und Zielgrößen

Das verkehrsstatistische Seminar der Deutschen Verkehrswissenschaftlichen Gesellschaft (DVWG) im Jahr 1988 bot eine exzellente Plattform für die Diskussionen alternativer Erhebungsansätze. Im Mittelpunkt des Interesses standen:

Anschrift des Verfassers:

Dipl.-Volksw. Günter Zimmermann
Bundesanstalt für Straßenwesen
Brüderstraße 53
5060 Bergisch Gladbach 1