

Die grafische Darstellung in der Korrespondenzanalyse

Gabler, Siegfried

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Gabler, S. (1993). Die grafische Darstellung in der Korrespondenzanalyse. *ZUMA Nachrichten*, 17(32), 22-37. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-209152>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Die grafische Darstellung in der Korrespondenzanalyse

Siegfried Gabler

Kontingenztafeln spielen in den Sozialwissenschaften eine bedeutende Rolle. Ein Ziel der Analyse solcher Kontingenztafeln ist dabei, Strukturen in den Daten zu erkennen sowie den Verlust bei der Projektion der Daten auf niedrigere Dimensionen angeben zu können. Wie übertragen sich überhaupt Strukturen in den Daten auf grafische Darstellungen? Und wie sieht und mißt man Abhängigkeiten zwischen Zeilen und Spalten?

Ein in den letzten Jahren verstärkt in den Vordergrund tretendes Verfahren zur grafischen Darstellung und Analyse zweidimensionaler Kontingenztafeln, ist die (einfache) Korrespondenzanalyse. Beim Lesen verschiedener Artikel zu diesem Thema entstand bei uns jedoch der Eindruck, daß die Grafiken oftmals nicht richtig interpretiert werden, weil unvergleichbare Punkte miteinander verglichen werden. Die folgende Übersicht soll daher klären, was bei der grafischen Darstellung von Kontingenztafeln in der Korrespondenzanalyse eigentlich gezeigt wird und welche Strukturen der Daten sich in den Grafiken widerspiegeln.

1. Ein kurzer Abriß der Korrespondenzanalyse

Ausgangspunkt einer Korrespondenzanalyse ist eine zweidimensionale Kontingenztafel N mit I Zeilen und J Spalten. Die Werte in der Kontingenztafel sind meist Häufigkeiten, also nichtnegativ und ganzzahlig. Das folgende einfache Demonstrationsbeispiel einer 6×3 Kontingenztafel wird uns im weiteren begleiten.

$$N = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 0 \\ 7 & 5 & 4 \end{pmatrix}$$

N_{ij} ist etwa die Anzahl der Betriebe des i -ten Wirtschaftszweiges in der j -ten Region ($i=1, \dots, 6; j=1, 2, 3$).

Wir gehen von der Häufigkeitstabelle N zur sogenannten *Korrespondenzmatrix*¹⁾ P über, indem wir N durch die Gesamtsumme N_{++} der Zellhäufigkeiten teilen. In unserem Beispiel ist $N_{++} = 32$.

Es sei $e_k = (0, \dots, 0, 1, 0, \dots, 0)'$ der k -te Einheitsvektor, das heißt, die k -te Komponente von e_k ist Eins, alle anderen Komponenten sind Null. e ist ein Vektor bestehend aus lauter Einsen. Die Länge der Vektoren ergibt sich aus dem jeweiligen Zusammenhang. Wir haben $N_{++} = e'Ne$. D_x bezeichne im weiteren stets die Diagonalmatrix mit x als Vektor der Diagonalelemente. In der Korrespondenzanalyse werden Punkte grafisch dargestellt, die in einem gewissen Zusammenhang mit den *Zeilenprofilen* bzw. *Spaltenprofilen* stehen.

Das i -te Zeilenprofil z_i ist definiert als i -te Zeile der Matrix $D_r^{-1}P$.

Das j -te Spaltenprofil s_j ist definiert als j -te Zeile der Matrix $D_c^{-1}P'$. Dabei ist $r = Pe$ der Vektor der Zeilensummen von P und $c = P'e$ der Vektor der Spaltensummen von P . Die Komponenten von z_i bzw. s_j lassen sich daher als relative Häufigkeiten interpretieren.

Im Beispiel ist $r' = (3, 5, 4, 3, 1, 16)/32$ und $c' = (14, 10, 8)/32$. Die Matrix der Zeilenprofile bzw. der Spaltenprofile lautet daher

$$D_r^{-1}P = \begin{pmatrix} 0.33 & 0.33 & 0.33 \\ 0.40 & 0.20 & 0.40 \\ 0.50 & 0.25 & 0.25 \\ 0.67 & 0.33 & 0 \\ 0 & 1 & 0 \\ 0.44 & 0.31 & 0.25 \end{pmatrix} \quad \text{bzw.} \quad D_c^{-1}P' = \begin{pmatrix} 0.07 & 0.1 & 0.13 \\ 0.14 & 0.1 & 0.25 \\ 0.14 & 0.1 & 0.13 \\ 0.14 & 0.1 & 0 \\ 0 & 0.1 & 0 \\ 0.50 & 0.5 & 0.5 \end{pmatrix}$$

Weiter ist

$$P - rc' = \begin{pmatrix} -0.0098 & 0.0020 & 0.0078 \\ -0.0059 & -0.0176 & 0.0234 \\ 0.0078 & -0.0078 & 0 \\ 0.0215 & 0.0020 & -0.0234 \\ -0.0137 & 0.0215 & -0.0078 \\ 0 & 0 & 0 \end{pmatrix}$$

$P=rc'$ würde *Unabhängigkeit* des Zeilen- vom Spaltenmerkmal bedeuten. Die Abweichungen $p_{ij}-r_jc_j$ von der Unabhängigkeit werden im bekannten χ^2 - Maß²⁾ zusammengefaßt. $\tau = \chi^2/N_{++}$ heißt *Trägheitsgewicht* oder *(Gesamt)Inertia*.

Definieren wir den χ^2 -Abstand zweier Vektoren x und y durch

$$d_q(x,y) = (x-y)' D_q^{-1} (x-y),$$

wobei q ein positiver Vektor ist, so läßt sich τ als gewichtete Summe³⁾ der Abstände der Zeilenprofile vom *Zeilenzentrum* c bzw. als gewichtete Summe der Abstände der Spaltenprofile vom *Spaltenzentrum* r auffassen. Im Beispiel ist $\tau = 0.1322$. Im Falle $q=e$ geht die χ^2 -Metrik in die übliche Euklidische Metrik über.

Wir sagen, daß ein Zeilenprofil $z'_i = e'_i D_r^{-1} P$ dem Zeilenprofil z'_j *näher* als dem Zeilenprofil z'_k ist, falls $d_c(z_i, z_j) < d_c(z_i, z_k)$ gilt.

Wir sagen, daß ein Spaltenprofil $s'_i = e'_i D_c^{-1} P$ dem Spaltenprofil s'_j *näher* als dem Spaltenprofil s'_k ist, falls $d_r(s_i, s_j) < d_r(s_i, s_k)$ gilt.

Mittels einer verallgemeinerten singulären Wertzerlegung können wir $P-rc'$ schreiben als $P-rc' = AD_\mu B'$ mit $A'D_r^{-1} A = id$ und $B'D_c^{-1} B = id$. id ist die Identitätsmatrix. μ ist der Vektor bestehend aus allen positiven Singulärwerten. Danach richtet sich auch die Zahl der Spalten von A und B . Wir definieren weiter:

$$F = D_r^{-1} A D_\mu, \quad G = D_c^{-1} B D_\mu \quad \text{und erhalten:}$$

$$D_r^{-1} (P-rc') = FB' \quad \text{und} \quad D_c^{-1} (P'-cr') = GA'.$$

$$\text{Weiter gilt } D_r^{-1} (P-rc') D_c^{-1} = F D_\mu^{-1} G'.$$

Im Beispiel ist

$$F = \begin{pmatrix} 0.030 & 0.230 \\ 0.317 & 0.182 \\ 0.110 & -0.097 \\ -0.236 & -0.562 \\ -1.406 & 0.473 \\ 0 & 0 \end{pmatrix} \quad \text{und} \quad G = \begin{pmatrix} 0.101 & -0.236 \\ -0.408 & 0.104 \\ 0.333 & 0.284 \end{pmatrix}$$

Die Matrix D_μ der positiven singulären Werte ist

$$D_\mu = \begin{pmatrix} 0.2904 & 0 \\ 0 & 0.2188 \end{pmatrix}$$

Der Vorteil der singulären Wertzerlegung ist, daß die χ^2 -Distanzen der Zeilen- bzw. Spaltenprofile zu Euklidischen Abständen der Zeilen von F bzw. G werden. Genauer gilt:

Satz 1.

- Der χ^2 -Abstand zwischen dem i -ten und j -ten Zeilenprofil ist gleich dem Euklidischen Abstand zwischen der i -ten und j -ten Zeile von F .
- Der χ^2 -Abstand zwischen dem i -ten und j -ten Spaltenprofil ist gleich dem Euklidischen Abstand zwischen der i -ten und j -ten Zeile von G .

Beweis:

a) Es ist

$$\begin{aligned} d_c(z_i, z_j) &= (e_i - e_j)' D_r^{-1} P D_c^{-1} P' D_r^{-1} (e_i - e_j) \\ &= (e_i - e_j)' D_r^{-1} (P - rc') D_c^{-1} (P' - cr') D_r^{-1} (e_i - e_j) \\ &= (e_i - e_j)' F B' D_c^{-1} B F' (e_i - e_j) \\ &= (e_i - e_j)' F F' (e_i - e_j). \end{aligned}$$

b) Es ist

$$\begin{aligned} d_r(s_i, s_j) &= (e_i - e_j)' D_c^{-1} P' D_r^{-1} P D_c^{-1} (e_i - e_j) \\ &= (e_i - e_j)' D_c^{-1} (P' - cr') D_r^{-1} (P - rc') D_c^{-1} (e_i - e_j) \\ &= (e_i - e_j)' G A' D_r^{-1} A G' (e_i - e_j) \\ &= (e_i - e_j)' G G' (e_i - e_j). \end{aligned}$$

Zwei weitere wichtige Formeln sind die sogenannten Übergangsformeln

$$F = D_r^{-1} P G D_\mu^{-1} \quad \text{und} \quad G = D_c^{-1} P' F D_\mu^{-1}.$$

Sie besagen zum einen, daß etwa die Berechnung der Koordinaten der Spaltenpunkte direkt aus den Zeilenpunkten und den singulären Werten erfolgen kann, zum andern, daß die Zeilenpunkte in der konvexen Hülle der Zeilen von $G D_\mu^{-1}$ liegen. Die Zeilenprofile liefern gerade die Gewichtungen der Eckpunkte. Für eine ausführliche Darstellung der Formeln in der Korrespondenzanalyse verweisen wir auf die Bücher von Benzécri (1992), Gifi (1990) Greenacre (1984) und Jambu (1992).

2. Grafische Darstellung

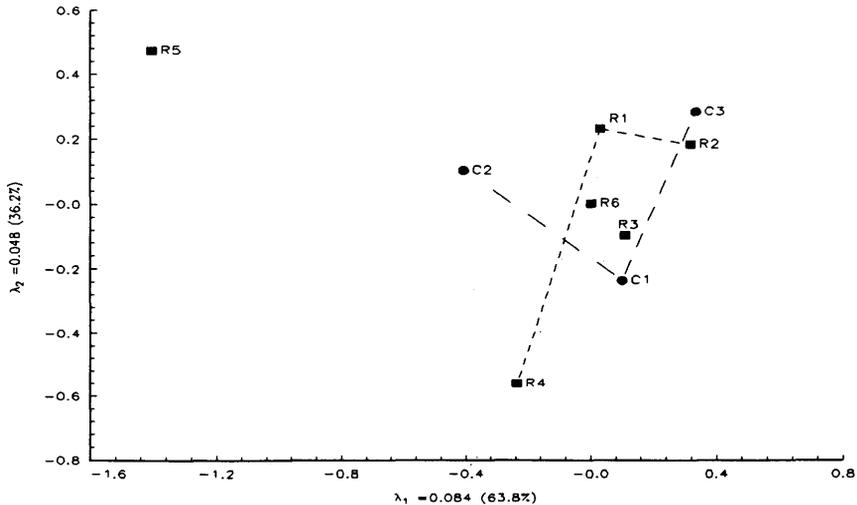
Was versteht man unter der grafischen Darstellung der Zeilen- und Spaltenprofile? In den meisten Fällen in der Literatur versteht man unter der (zweidimensionalen) grafischen Darstellung der Zeilen- und/oder Spaltenprofile das Eintragen der Zeilen der Koordinatenmatrizen F und/oder G (bzw. deren ersten beiden Spalten) in ein Euklidisches Koordinatensystem. In F (G) stehen die sogenannten *Prinzipalkoordinaten* der Zeilenprofile (Spaltenprofile). Man identifiziert in diesem Sinne das i -te Zeilenprofil (Spaltenprofil) mit der i -ten Zeile von F (G). Wir nennen in der Grafik die Zeilen von F (G) *Zeilenpunkte* (*Spaltenpunkte*). Die Gesamtheit aller Zeilenpunkte (Spaltenpunkte) heißt *Zeilenpunktwolke* (*Spaltenpunktwolke*). Den Vektor vom Ursprung zum Zeilenpunkt (Spaltenpunkt) nennen wir *Zeilenpunktvektor* (*Spaltenpunktvektor*). Unter die Achsen werden die singulären Werte bzw. der prozentuale Anteil der Hauptachsen an der Inertia geschrieben. Eine gute Repräsentierung der Profile in der Grafik hat man natürlich nur, wenn die beiden ersten Achsen einen großen Anteil an der Inertia haben. Nur dann ist eine Interpretation über Distanzen gerechtfertigt. Die Nähe von Zeilenprofilen im Datensatz wird mit Hilfe der χ^2 -Distanz gemessen: Nach dem im Satz des letzten Kapitels Gezeigten entsprechen diese Distanzen den Euklidischen Abständen der Zeilenpunkte, d.h. der entsprechenden Zeilen der F Matrix. Analoges gilt für die Spaltenprofile. Werden die Einheiten auf beiden Achsen gleich lang gewählt, so lassen sich in der Grafik die Abstände durch ein Lineal messen. Profile werden so für das Auge vergleichbar.

Im folgenden werden acht Strukturmerksätze aufgeschrieben, die es dem Nutzer erleichtern sollen, Grafiken in der Korrespondenzanalyse richtig zu lesen. Dabei steht im Vordergrund, Strukturen in den Daten in der Grafik wiederzufinden.

Strukturmerksatz 1: Ähnliche Zeilenprofile (Spaltenprofile) entsprechen in der grafischen Darstellung Zeilenpunkten, die nahe beieinander liegen. Sehr unterschiedliche Zeilenprofile (Spaltenprofile) entsprechen in der grafischen Darstellung Zeilenpunkten (Spaltenpunkte), die weit auseinander liegen.

Die Abstände zwischen Zeilen- und Spaltenpunkten in der Abbildung lassen sich nicht als χ^2 -Distanzen der entsprechenden Profile interpretieren, obwohl dies oft gemacht wird!

Abbildung 1



Im Beispiel (Abbildung 1) ist der Zeilenpunkt R1 näher bei R2 als bei R4. Das erste Zeilenprofil ist daher dem zweiten Zeilenprofil ähnlicher als dem vierten Zeilenprofil. Der Spaltenpunkt C1 ist dem Spaltenpunkt C3 (euklidisch) kaum näher als dem Spaltenpunkt C2. Das erste Spaltenprofil ist daher dem dritten Spaltenprofil kaum näher als dem zweiten. Eine Aussage wie: "Das erste Zeilenprofil ist dem zweiten Spaltenprofil näher als das fünfte Zeilenprofil" ist inhaltlich und methodisch falsch.

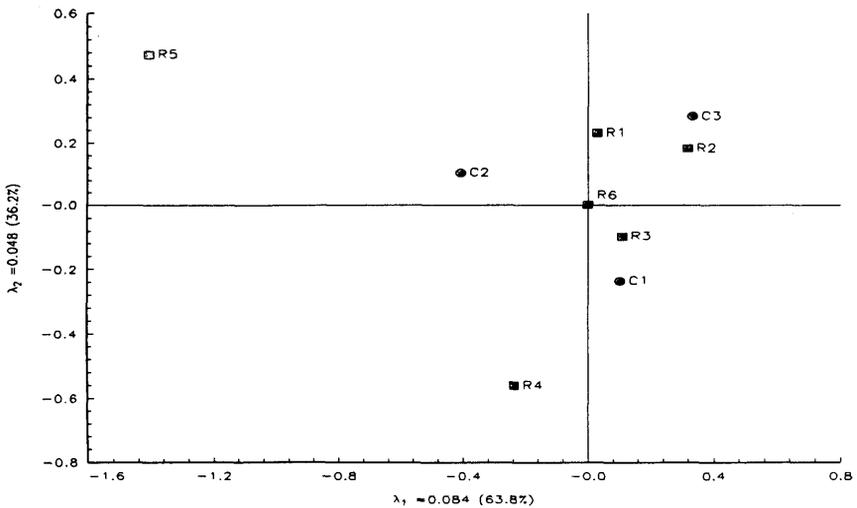
Strukturmerksatz 2: Der Zentroid c' der Zeilenprofile entspricht in der Abbildung dem Ursprung. Analog kommt der Zentroid r' der Spaltenprofile in der Abbildung im Ursprung zu liegen.

Beweis:

Ist $e_1'D_R^{-1}P = c'$, so gilt $e_1'D_R^{-1}(P-rc') = 0$ und daher $e_1'FB' = 0$.

Multiplikation von rechts mit $D_C^{-1}B$ liefert $e_1'F = 0$. Der Beweis für r' verläuft analog.

Abbildung 2



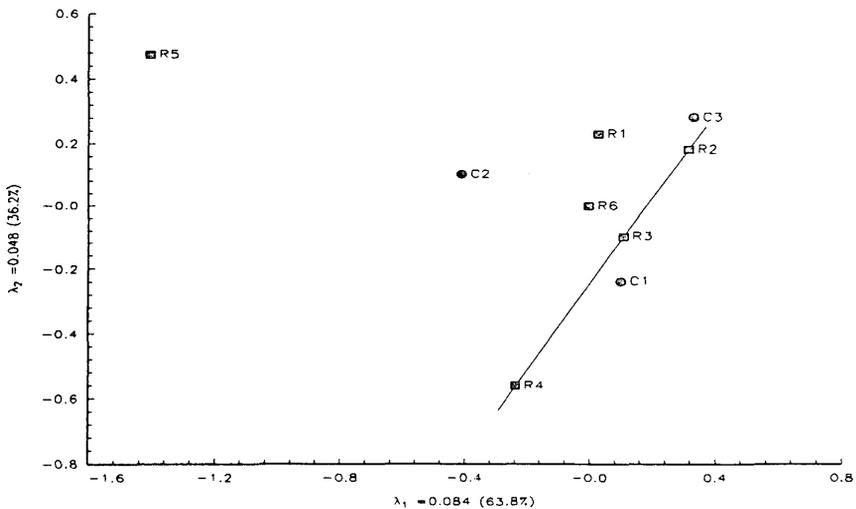
Im Beispiel (Abbildung 2) ist das sechste Zeilenprofil identisch mit c' . Daher liegt R6 in der Abbildung im Ursprung.

Strukturmerksatz 3: Die i -, j - und k -ten Zeilenprofile liegen im Datensatz genau dann auf einer Geraden, wenn die entsprechenden Zeilenpunkte in der Abbildung auf einer Geraden liegen. Die Proportionen bleiben dabei erhalten. Analoges gilt bei Spaltenprofilen.

Beweis: Es ist

$0 = (e_k - \lambda e_i - (1-\lambda)e_j)' D_r^{-1} P = (e_k - \lambda e_i - (1-\lambda)e_j)' D_r^{-1} (P - rc') = (e_k - \lambda e_i - (1-\lambda)e_j)' FB'$ und daher $(e_k - \lambda e_i - (1-\lambda)e_j)' F = 0$. Der Beweis bei den Spaltenprofilen verläuft analog.

Abbildung 3



Im Beispiel (Abbildung 3) liegen R2, R3 und R4 auf einer Geraden, weil das zweite, dritte und vierte Zeilenprofil auf einer Geraden liegen. Genauer ist $R3 = 5/8 \cdot R2 + 3/8 \cdot R4$.

Wenn es um Vergleiche zwischen Zeilen- und Spaltenprofilen geht, ist die Prinzipal-darstellung der Profile, also das Zeichnen der Zeilen von F und G, in der Abbildung

nicht geeignet, was oft mißachtet wird (vgl. Sawatzke 1991). In Abbildung 3 sind R5 und C2 euklidisch weit auseinander, obwohl ein starker Zusammenhang besteht wie in Abbildung 7 zu sehen sein wird. Bei Vergleichen von Zeilen- mit Spaltenpunkten sollten in der grafischen Darstellung die Zeilen von $FD_{\mu}^{-\alpha}$ und $GD_{\mu}^{-(1-\alpha)}$ für $0 \leq \alpha \leq 1$ gewählt werden. Das Skalarprodukt der beiden Matrizen ist offensichtlich von α unabhängig. Wir nennen die grafische Darstellung α -Darstellung, sie heißt im Falle $\alpha \neq 0.5$ *asymmetrisch*. Insbesondere der Fall $\alpha=0$ ist wichtig. Die Zeilen von GD_{μ}^{-1} heißen *Spaltenpunkte* in der *Standarddarstellung*. Die Zeilenprofile werden weiter in den Prinzipalkoordinaten dargestellt. Analog ist der Fall $\alpha=1$ zu sehen.

Einen wichtigen Aspekt in der Korrespondenzanalyse bilden die *supplementären Zeilenprofile* bzw. *Spaltenprofile*. In welchen Punkt wird irgendein supplementäres Zeilenprofil f' in der Grafik abgebildet?

Aus $(f-c)' = x'B'$ und $G = D_c^{-1}BD_{\mu}$ folgt $x' = (f-c)'D_c^{-1}B = (f-c)'GD_{\mu}^{-1} = f'GD_{\mu}^{-1}$, weil $c'G = 0$ gilt.

Also ist das Zeichnen der Spaltenpunkte in der Standarddarstellung identisch mit dem Einzeichnen der Einheitsmatrix als supplementäre Zeilenprofile in die Prinzipaldarstellung (der Zeilenpunkte). Daher lassen sich auf diesem Weg Distanzen zwischen Zeilen- und Spaltenpunkten interpretieren.

Da alle singulären Werte μ_i größer Null sind, gelten die Strukturmerksätze 1 bis 3 auch für $FD_{\mu}^{-\alpha}$ statt für F.

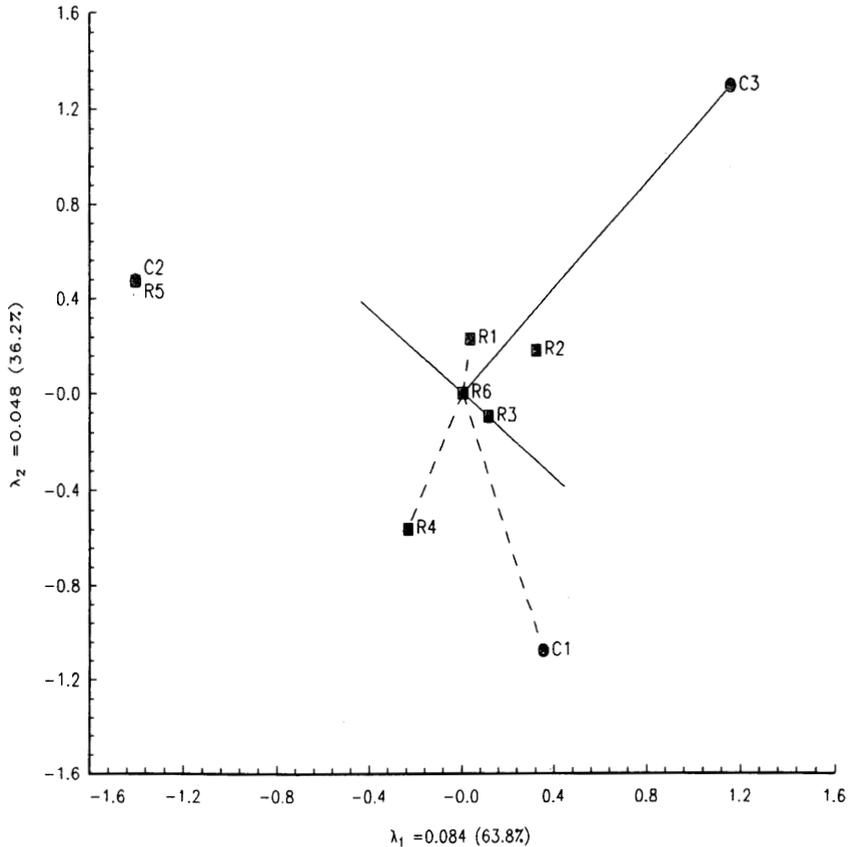
Wenn in der Abbildung die Einheit auf der ersten (euklidischen) Achse genauso lang ist wie auf der zweiten Achse lassen sich (normierte) Skalarprodukte, geometrisch Winkel, einfach interpretieren. Die nachfolgenden Strukturmerksätze 4, 5 und 6 gelten für beliebiges α , die Sätze 7 und 8 im Falle $\alpha=0$, das heißt der Standarddarstellung. Für die Abbildungen wurde in allen Fällen $\alpha=0$ gewählt.

Strukturmerksatz 4: Es ist $p_{ij} > (=, <) r_i c_j$ gleichbedeutend mit spitzem (rechtem, stumpfem) Winkel zwischen i -tem Zeilenpunkt und j -tem Spaltenpunkt in der α -Darstellung.

Beweis: Wegen $D_r^{-1}(P-rc')D_c^{-1} = FD_\mu^{-1}G'$ ist

$$\frac{p_{ij}}{r_i c_j} - 1 = e_i' D_r^{-1}(P-rc')D_c^{-1} e_j = e_i' F D_\mu^{-1} G' e_j = e_i' F D_\mu^{-\alpha} D_\mu^{-(1-\alpha)} G' e_j.$$

Abbildung 4



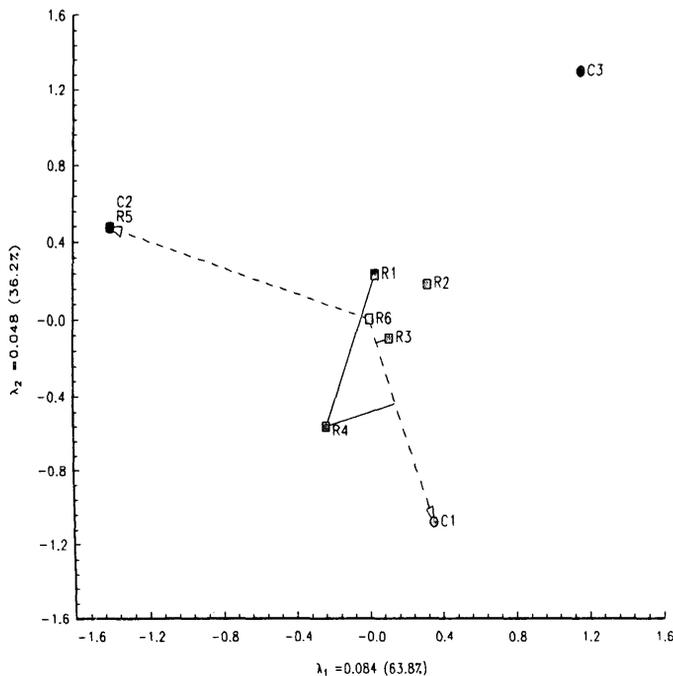
Im Beispiel (Abbildung 4) stehen der dritte Zeilenpunktvektor und der dritte Spaltenpunktvektor senkrecht aufeinander. In den Daten muß daher $p_{33} = r_3 c_3$ gelten. Der

spitze Winkel zwischen dem vierten Zeilenpunktvektor und dem ersten Spaltenpunktvektor liefert $p_{41} > r_4 c_1$. Der stumpfe Winkel zwischen dem ersten Zeilenpunktvektor und dem ersten Spaltenpunktvektor liefert $p_{11} < r_1 c_1$.

Strukturmerksatz 5: Die k -te Komponente des i -ten Zeilenprofils ist genau dann kleiner als die k -te Komponente des j -ten Zeilenprofils, wenn die Projektion des i -ten Zeilenpunktes auf den k -ten Spaltenpunktvektor kleiner ist als die des j -ten Zeilenpunktes. Fallen die Projektionen von Zeilenpunkten auf einen Spaltenpunktvektor zusammen, steht also der Spaltenpunktvektor orthogonal zur Verbindungsgeraden von Zeilenpunkten, so sind die Komponenten der entsprechenden Zeilenprofile in der entsprechenden Spalte gleich. Beweis: Es ist

$$(e_i - e_j)' F D_{\mu}^{-1} G' e_k = (e_i - e_j)' D_r^{-1} (P - r c') D_c^{-1} e_k = (e_i - e_j)' D_r^{-1} P D_c^{-1} e_k = \left(\frac{p_{ik}}{r_i} - \frac{p_{jk}}{r_j} \right) \frac{1}{c_k}$$

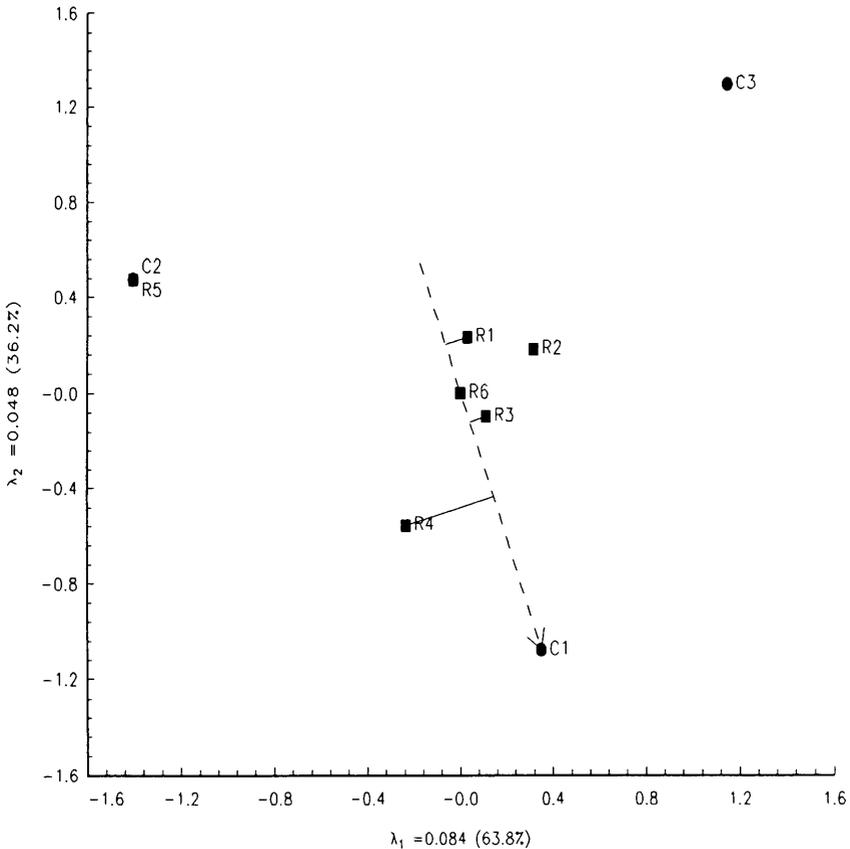
Abbildung 5



Im Beispiel (Abbildung 5) ist die Projektion des vierten Zeilenpunktes auf den ersten Spaltenpunktvektor länger als die Projektion des dritten Zeilenpunktes. Daher ist $p_{41}/r_4 > p_{31}/r_3$. Da der Spaltenpunktvektor zu C2 orthogonal zur Verbindungsgeraden der Zeilenpunkte R1 und R4 ist, gilt folgendes: Das Verhältnis von p_{12} und p_{42} ist gleich dem Verhältnis von r_1 und r_4 .

Strukturmerksatz 6: Die Verhältnisse der Abstände von Projektionen der Zeilenpunkte auf den k -ten Spaltenpunktvektor sind dieselben wie die Verhältnisse der Abstände der k -ten Komponenten der Zeilenprofile.

Abbildung 6



$$\text{Beweis: Es ist } \frac{(e_i - e_j)' FD_\mu^{-1} G' e_k / e_k' GD_\mu^{-2(1-\alpha)} G' e_k}{(e_i - e_j)' FD_\mu^{-1} G' e_k / e_k' GD_\mu^{-2(1-\alpha)} G' e_k} = \frac{\frac{P_{ik}}{r_i} - \frac{P_{jk}}{r_j}}{\frac{P_{ik}}{r_i} - \frac{P_{jk}}{r_j}}$$

Im Beispiel (Abbildung 6) folgt beispielsweise aus der Tatsache, daß die Projektion von R3 auf den Spaltenpunktvektor zu C1 in der Mitte der Projektionen von R1 und R4 liegt, daß $p_{31}/r_3 = 1/2(p_{11}/r_1 + p_{41}/r_4)$ gilt.

Strukturmerksatz 7: *Es sei $I \geq J$. In der Standarddarstellung fällt genau dann der i -te Zeilenpunkt mit dem j -ten Spaltenpunkt zusammen, wenn die i -te Zeile der Korrespondenzmatrix nur in der j -ten Spalte von Null verschieden ist.*

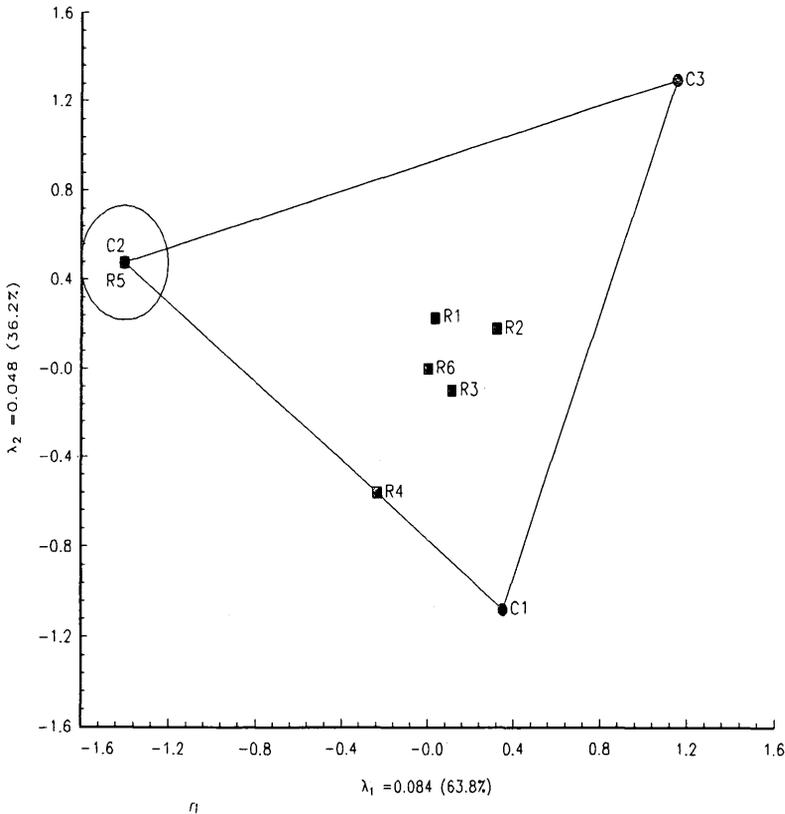
Beweis: Es ist

$$e_i' F = e_i' (D_r^{-1} P) G D_\mu^{-1} = e_j' G D_\mu^{-1}.$$

Also muß auch $(e_i' D_r^{-1} P - e_j') G = 0$ sein, was wegen der Unabhängigkeit der Zeilen von G zur Folge hat: $e_i' D_r^{-1} P = e_j'$. Nur die j -te Komponente der i -ten Zeile von $D_r^{-1} P$ ist daher von Null verschieden und hat den Wert Eins. Dies ist nur möglich, falls $p_{ij} = r_i$ und $p_{ik} = 0$ für $k \neq j$ gilt.

Im Beispiel (Abbildung 7) sind alle Zeilenpunkte Konvexkombinationen der Spaltenpunkte C1, C2, C3 in der Standarddarstellung. Da R5 und C2 zusammenfallen, steht in der fünften Zeile der Zeilenprofilmatrix nur in der zweiten Komponente eine Eins; alle anderen Komponenten sind Null. Das fünfte Zeilenprofil lädt daher nur auf die zweite Spalte.

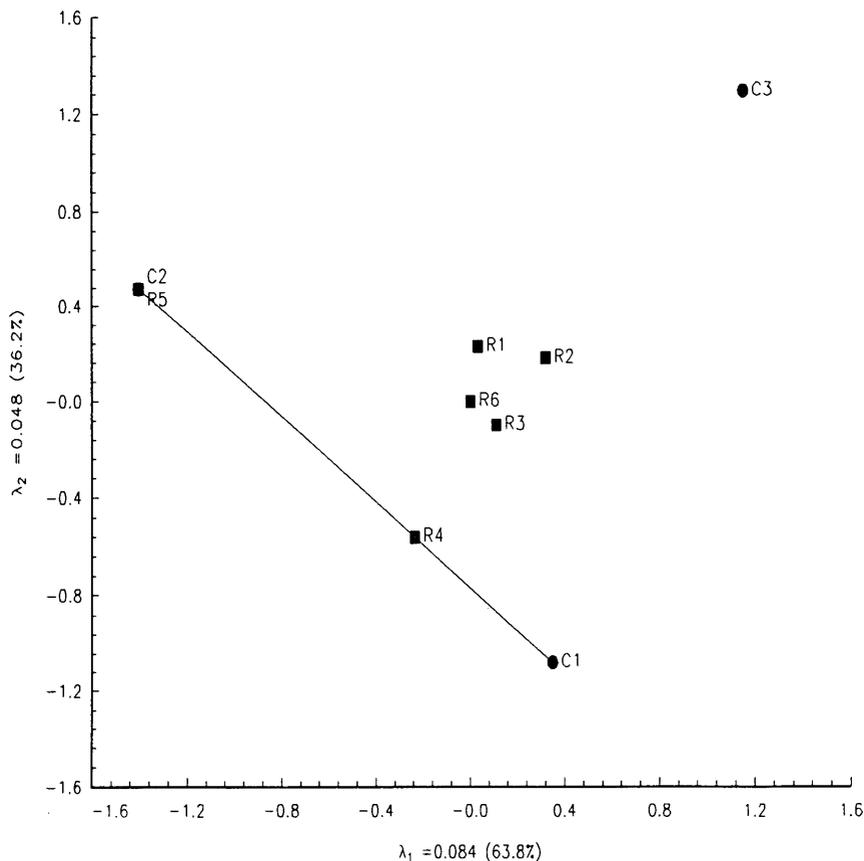
Abbildung 7



Strukturmerksatz 8 (Verallgemeinerung von Strukturmerksatz 7): Es sei $I \geq J$. In der Standarddarstellung liegt der i -te Zeilenpunkt genau dann auf der Verbindungsgeraden zwischen dem j -ten und k -ten Spaltenpunkt, wenn die i -te Zeile der Korrespondenzmatrix nur in der j -ten und k -ten Spalte von Null verschieden ist. Die Konvexkombination der Spaltenpunkte entspricht in der Abbildung den j -ten und k -ten Komponenten des i -ten Zeilenprofils.

Beweis: Analog Strukturmerksatz 7.

Abbildung 8



Beispiel (Abbildung 8) liegt R4 auf der Verbindungsgeraden zwischen C1 und C2. Die entsprechende vierte Zeile der Korrespondenzmatrix hat also nur in der ersten und zweiten Spalte von Null verschiedene Komponenten.

Da $R_4 = 2/3 C_1 + 1/3 C_2$ gilt ist $p_{41} / r_4 = 2/3$ und $p_{42} / r_4 = 1/3$.

In der Standarddarstellung werden allgemein Zeilenprofile, bei denen einige Komponenten null sind, in Zeilenpunkte abgebildet, die Konvexkombinationen nur der entsprechenden Spaltenpunkte sind. Es bleibt noch anzumerken, daß bei unserem einfachen Beispiel die zweidimensionale grafische Darstellung die Zeilenwolke

(Spaltenwolke) zu 100% darzustellen vermag, da die Datenmatrix nur drei Spalten hat. Im allgemeinen wird die Repräsentierung der Profile durch die zweidimensionale Punkt wolke nicht 100%ig sein. In der Regel sind die ad hoc Schlüsse mit den Augen aus der grafischen Darstellung dann nur mehr oder weniger approximativ richtig. Der Unterschied zwischen der Prinzipal- und Standarddarstellung besteht nur darin, daß die Koordinaten der Spaltenpunkte durch die der Größe nach geordneten positiven singulären Werte dividiert werden. Die Spaltenpunkte bleiben daher in der Prinzipal- und Standarddarstellung im gleichen Quadranten. Sie wandern praktisch von der Prinzipaldarstellung mehr oder weniger in Richtung der Achsen mit den kleinsten singulären Werten nach außen. Ein Experte kann daher auch schon an Hand der Prinzipaldarstellung einiges über den Zusammenhang zwischen Zeilen- und Spaltenprofilen aussagen. Ein häufiger Nachteil bei der grafischen Darstellung in der Standarddarstellung der Spaltenprofile ist die Klumpung der Zeilenpunkte um den Ursprung. Die Häufigkeiten sind dann in der Datentabelle über die Spalten verteilt. Dominante Spalten gibt es dann nicht. Was hier im Zusammenhang mit der Standarddarstellung der Spalten gesagt wurde, gilt analog im Falle, daß die Standarddarstellung für die Zeilenprofile und die Prinzipaldarstellung für die Spaltenprofile gewählt wird.

Anmerkungen

$$1) P = \frac{N}{N_{++}} \quad \text{mit} \quad N_{++} = \sum_{i=1}^I \sum_{j=1}^J N_{ij}$$

$$2) \chi^2 = N_{++} \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - r_i c_j)^2}{r_i c_j}$$

$$3) \tau = \sum_i r_i \sum_j \frac{1}{c_j} \left(\frac{p_{ij}}{r_i} - c_j \right)^2 = \sum_i r_i d_c(z_i, c) = \sum_j c_j d_r(s_i, r) = \sum_j c_j \sum_i \frac{1}{r_i} \left(\frac{p_{ij}}{c_j} - r_i \right)^2$$

Literatur

- Benzécri, J.-P., 1992: *Correspondence Analysis Handbook*. New York: Marcell Dekker.
- Gifi, A., 1990: *Nonlinear Multivariate Analysis*. Chichester: John Wiley & Sons.
- Greenacre, M.J., 1984: *Theory and Applications of Correspondence Analysis*. Academic Press New York.
- Jambu, M., 1992: *Explorative Datenanalyse*. Gustav Fischer Stuttgart.
- Sawatzke, F., 1991: MDS, Correspondence Analysis und Biplot: drei Verfahren zur räumlichen Darstellung von Kreuztabellen. *Planung und Analyse* 1991: 89-92.