Kanonische Korrelation[1]

Evaluation und Forschungsmethoden

U. Mortensen

FB Psychologie und Sportwissenschaften, Institut III
Westfälische Wilhelms-Universität Münster

Inhaltsverzeichnis

[1]Revidiert: 28. 05. 2005

1 Einführung

Viele psychologische Fragestellungen führen auf die Notwendigkeit, die Kovariationen bzw. Korrelationen zwischen Variablen interpretieren zu müssen. So ist etwa nach dem Zusammenhang zwischen *einer* bestimmten Variable Y und den Variablen $X_1, \ldots, X_p$ gefragt: die Analyse dieses Zusammenhanges führt auf die multiple Korrelation; Y ist dann die *Kriteriumsvariable*, und die X_j sind die *Prädiktorvariablen.* Diese Frage stellt sich aber oft in allgemeinerer Form, wenn nicht nur eine, sondern mehrere Kriteriumsvariablen $Y_1, \ldots, Y_q$, $q > 1$, gegeben sind:

1. Es wird vermutet, dass zwischen den Persönlichkeitsvariablen $X_1, \ldots, X_p$ und den physiologischen Variablen $Y_1, \ldots, Y_q$ ein Zusammenhang besteht; lassen sich die $X_1, \ldots, X_p$ aufgrund der $Y_1, \ldots, Y_q$ vorhersagen, oder kann man umgekehrt von den $Y_1, \ldots, Y_q$ auf die $X_1, \ldots, X_p$ schließen?

2. Es werden m Persönlichkeitsmerkmale in zwei verschiedenen sozialen Situationen gemessen; in der Situation S_1 seien $X_1, \ldots, X_p$ die Maße für diese Variablen, und in der Situation S_2 seien $Y_1, \ldots, Y_q$ die Maße. Sind die Maße situationsunabhängig, so sollten die Messungen Y_j aufgrund der Messungen X_j gut vorhersagbar sein.

3. Die Variablen $V_1, \ldots, V_n$ werden vor und nach einer "Behandlung" (Therapie, Veränderung der Arbeitssituation oder der experimentellen Anordnung) gemessen, wobei X_j, $j = 1, \ldots, n$ die Messungen der V_j vor, und die Y_j die Messungen der V_j nach der Behandlung seien. Wieder ist die Frage, ob die Y_j aufgrund der X_j "vorausgesagt" werden können.

4. Es wird ein Fragebogen zur Evaluation einer Lehrveranstaltung vorgelegt. Der Fragebogen enthält insgesamt n Items; n_1 der Items beziehen sich auf Merkmale der Studierenden (Vorwissen, Interesse an den Inhalten der Veranstaltung), n_2 auf Merkmale des Dozenten (Vortragsstil und Sprache, Bereitschaft, auf Fragen einzugehen, etc). Die Frage ist, ob die Antworten auf diese n_2 Items durch die Antworten auf die n_1 Items, die die Studierenden charakterisieren, vorhergesagt werden können.

Vorbemerkung: Bei den folgenden Betrachtungen wird voraugesetzt, dass alle Variablen *standardisiert* sind, d.h.für die Komponenten x_{ij} des j-ten Vektors X_j mußte eigentlich $z_{ij}(x) = (x_{ij} - \bar{x}_j)/s_j$ geschrieben; analog dazu $z_{ij}(y) = (y_{ij} - \bar{y}_j)/s_j$. Um eine zu komplizierte Notation zu vermeiden ($z_{ij}(x)$ und $z_{ij}(y)$, etc, werden diese z-Werte wieder in x_{ij} und y_{ij} umbenannt. Die Standardisierung ist nicht unbedingt notwendig, wenn alle Variablen in gleichen Einheiten gemessen werden. Man vermeidet aber Fragen der Vergleichbarkeit von Maßeinheiten, wenn man von standardisierten Werten ausgeht.[2] □

Man könnte nun jede Variable Y_j anhand der $X_1, \ldots, X_n$ durch Anwendung der multiplen Korrelation vorhersagen. Die Y_j - ebenso wie die X_j werden aber im allgemeinen untereinander korreliert sein, so dass man ein ebenso redundantes wie schwer zu interpretierendes System von Regressionsgleichungen bekäme. Andererseits ist aus der Theorie der PCA bekannt, dass die X_j, $j = 1, \ldots, p$ und die Y_k, $k = 1, \ldots, q$, jeweils als Linarkombinationen unabhängiger, insbesondere orthogonaler Vektoren dargestellt werden können, die als

[2] Mit dem Zeichen □ wird das Ende einer hervorgehobenen Bemerkung oder Anmerkung, eines Beispiels oder eines Beweises bezeichnet.

"latente" Dimensionen interpretiert werden können. Es sei $X = [X_1, \ldots, X_p]$ die Matrix, die aus den Spaltenvektoren X_j gebildet werden kann, und $Y = [Y_1, \ldots, Y_q]$ die entsprechenende Matrix der Y_k-Vektoren. Man kann stets $X = L_x P_x'$ und $Y = L_y P_y'$ ansetzen, wobei L_x und L_y die Matrizen mit den orthogonalen Basisvektoren (als Spaltenvektoren) sind und P_x, P_y die Matrizen der "Gewichte", die zur Darstellung der individuellen Variablenvektoren X_j bzw. Y_k benötigt werden. Die Matrizen P_x und P_y sind orthonormal, so dass man auch die latenten Vektoren in L_x und L_y als Linearkombinationen der Vektoren $X_j$bzw. Y_k benützen kann: man hat dann $L_x = XP_y$, $L_y = YP_y$. Die latenten Variablen in L_x und L_y werden dann wie Kriteriumsvariablen "vorausgesagt". Die Frage nach der Beziehung zwischen den Variablen in X und Y kann dann auch als Frage nach der Beziehung zwischen L_x und L_y gestellt werden. Sollte $L_x = L_y$ gelten, so kann man sagen, dass sich die X_j und Y_k als Linearkombinationen derselben latenten Dimensionen darstellen lassen. $L_x = L_y$ impliziert dann $XP_x = YP_y$, und wegen der Orthonormalität von P_y und P_x erhält man daraus sofort die Beziehungen $Y = XP_xP_y'$ bzw. $X = YP_yP_x'$, d.h. X und Y lassen sich perfekt auseinander vorhersagen!

Der Idealfall $L_x = L_y$ wird in den seltensten Fällen vorliegen, er illustriert aber, wie im Prinzip die Beziehung zwischen X- und Y-Variablen über die jeweiligen latenten Variablen charakterisiert werden kann. Typischerweise wird $L_x \neq L_y$ gelten. Man kann aber fragen, worin bzw. wie weit sich die Vektoren in L_x und L_y unterscheiden. Diese Frage ist äquivalent der Frage, ob sich die Vektoren in L_x und L_y so rotieren lassen, dass die Korrelation zwischen bestimmten Paaren (L_1^x, L_1^y), (L_2^x, L_2^y) etc maximal ist. L_j^x und L_j^y sind dabei rotierte Basisvektoren für X bzw. Y, und (L_1^x, L_1^y) ist das Paar rotierter Basisvektoren, deren Korrelatio maximal ist, (L_2^x, L_2^y) ist das Paar mit der zweithöchsten Korrelation, etc. Je höher diese Korrelationen sind, desto eher kann man sagen, dass die Variablen in X und Y ähnliche latente Strukuren haben.

2 Kanonische Variablen und ihre Bestimmung

2.1 Definition der Kanonischen Variablen

Man könnte also die Frage nach der Beziehung zwischen den X- und Y-Variablen angehen, indem man die latenten Vektoren L_x und L_y bestimmt und dann versucht, diejenige Rotation zu finden, für die die eben genannten maximalen Korrelationen gelten. Es zeigt sich aber, dass ein anderer Weg direkter zum Ziel führt: man bestimmt die latenten Vektoren von vornherein so, dass die Korrelationen die jeweils maximal möglichen sind:

Annahme: Es existieren orthogonale Vektoren $\vec{U}_1, \ldots, \vec{U}_r$ sowie orthogonale Vektoren $\vec{V}_1, \ldots, \vec{V}_r$ derart, dass einerseits die Beziehungen

$$\vec{U}_s = a_{s1}\vec{X}_1 + a_{s2}\vec{X}_2 + \cdots + a_{sp}\vec{X}_p, \tag{1}$$

$$\vec{V}_s = b_{s1}\vec{Y}_1 + b_{s2}\vec{Y}_2 + \cdots + b_{sq}\vec{Y}_q, \quad s = 1, \ldots, r \tag{2}$$

gelten, und dass andererseits

$$r(\vec{U}_1, \vec{V}_1) \geq r(\vec{U}_2, \vec{V}_2) \geq \cdots \geq r(\vec{U}_s, \vec{V}_s) \tag{3}$$

gilt, wobei $r(\vec{U}_1, \vec{V}_1)$ die maximal mögliche Korrelation sein soll, $r(\vec{U}_2, \vec{V}_2)$ die maximal mögliche Korrelation nach $r(\vec{U}_1, \vec{V}_1)$ ist, usw. □

Schreibweisen: Der Einfachheit halber wird gelegentlich R_{uv} statt $r(U_j, V_j)$, für irgend ein j, $1 \leq j \leq r$ geschrieben. Für jede der p Variablen $X_1, \ldots, X_p$ und jede der q Variablen $Y_1, \ldots, Y_q$ liegen Messungen von N Personen (allgemein: Objekten) vor. Für die i-te Person hat man also die Gleichungen

$$\begin{aligned} u_{is} &= a_{s1}X_{i1} + \cdots + a_{sp}X_{ip} & (4)\\ v_{is} &= b_{s1}Y_{i1} + \cdots + b_{sq}Y_{iq}, \quad i = 1, \ldots, N, \quad s = 1, \ldots, r & (5)\end{aligned}$$

In Matrixform kann man für U_s und V_s die Gleichungen

$$\begin{aligned} \vec{U}_s &= X\vec{a}_s, \quad \vec{a}_s = (a_{s1} \ldots, a_{sp})' & (6)\\ \vec{V}_s &= Y\vec{b}_s, \quad \vec{b}_s = (b_{s1}, \ldots, b_{sq})' & (7)\end{aligned}$$

schreiben. Die Komponenten der Vektoren $\vec{a}_s$ und $\vec{b}_s$ können als Regressionsgewichte aufgefaßt werden. Man kann dann die Variablen U_s und V_s, $s = 1, \ldots, r$ für $r = \min(p, q)$ zu (n, q)-Matrizen U und V zusammenfassen, und die zugehörigen Gewichtsvektoren $\vec{a}_1, \ldots, \vec{a}_r$ und $\vec{b}_1, \ldots, \vec{b}_r$ zu (r, r)-Matrizen $\vec{a}$ und $\vec{b}$. Die Gleichungen (6) und (7) können dann zu

$$\begin{aligned} U &= XA', & (8)\\ V &= YB', & (9)\end{aligned}$$

verallgemeinert werden.

Bevor auf die inhaltliche Bedeutung der Annahme eingegangen wird, soll noch eine Definition vorangestellt werden:

Definition 1 *Für den Fall, dass $\vec{a}$ und $\vec{b}$ die Korrelation R_{uv} maximieren, heißt R_{uv}* kanonische Korrelation, *wobei $R_{uv} \geq 0$ gesetzt wird. Die Vektoren $\vec{U}_s$ und $\vec{V}_s$ heißen* kanonische Faktoren, *oder* kanonische Variablen. *Die Komponenten bzw. Elemente von $\vec{a}_s$ und $\vec{b}_s$ enthalten die* kanonischen Gewichte *der Variablen.*

Anmerkungen:

1. Die Einschränkung auf positive R_{uv}-Werte ist deshalb möglich, weil die Vektoren $\vec{a}$ und $\vec{b}$ stets so gewählt werden können, dass diese Bedingung erfüllt ist.

2. Die Vektoren von U und V entsprechen rotierten latenten Vektoren für X bzw. Y. Sie haben so viele Komponenten, wie es Personen gibt; die Komponenten *entsprechen* also den Faktorscores in der PCA, sind aber nicht notwendig gleich den Faktorscores.

3. Die Anzahl von Dimensionen, die sowohl für die Menge S_1 als auch für die Menge S_2 betrachtet werden können, ist maximal[3] gleich $\min(p, q)$, wenn p die Anzahl der Variablen in S_1 und q die Anzahl der Variablen in S_2 ist, so dass $\max r = \min(p, q)$ gilt. Ist z.B. $p = \min(p, q)$, so kann 100 % der Varianz der Variablen in S_1 erklärt werden, und i.a. wird weniger Varianz der Variablen S_2 erklärt.

[3] max steht für "maximal" oder "der Maximalwert von ...", min für "minimal" oder "das Minimum von ...".

Es muß noch geklärt werden, was unter *maximaler Korrelation* verstanden werden soll. Die Korrelation ist als Produkt-Moment-Korrelation definiert:

$$r(\vec{U}_s, \vec{V}_s) = \frac{Kov(\vec{U}_s, \vec{V}_s)}{\sqrt{Var(\vec{U}_s) Var \vec{V}_s)}}, \quad s = 1, \ldots, r,$$

und dies bedeutet $-1 \leq r(\vec{U}_s, \vec{V}_s) \leq 1$. Würde man die Gewichte $a_{s1}, \ldots, a_{sp}$ und $b_{s1}, \ldots, b_{sq}$ völlig frei wählen können, so könnte man stets die maximale Korrelation $r(\vec{U}_s, \vec{V}_s) = 1$ erreichen. Eine beliebige Wahl der Gewichte ist aber uninteressant, weil die $\vec{U}_s$, $\vec{V}_s$ dann nicht mehr notwendig latente Dimensionen abbilden; die Wahl der Gewichte muß eben so erfolgen, *dass* latente Dimensionen abgebildet werden. Dies bedeutet, dass bei der Wahl der Gewichte bestimmte Strukturen in den Daten X und Y berücksichtigt werden müssen. Aus (1) und (2) folgt nun, dass die Varianz der $\vec{U}_s$, $\vec{V}_s$ durch die Varianz der $\vec{X}_s$ bzw. $\vec{Y}_s$ bestimmt ist; (1) impliziert

$$Var(\vec{U}_s) = Var(a_{s1}X_1 + a_{s2}X_2 + \cdots + a_{sp}X_p),$$

und eine analoge Aussage gilt für $Var(\vec{V}_s)$. Nun kann man die Vektoren $\vec{U}_s$ und $\vec{V}_s$ stets standardisieren, so dass

$$\vec{U}_s{}'\vec{U}_s = \vec{V}_s' \vec{V}_s = 1 \tag{10}$$

gilt. Dies bedeutet ja nur, dass die Länge der Vektoren gleich 1 gesetzt wird, die *relative* Größe der individuellen Komponenten, d.h. die Größenverhältnisse der Komponenten von $\vec{U}_s$ und $\vec{V}_j$ wird dadurch nicht verändert. Es zeigt sich, dass (10) eine hinreichende Nebenbedingung für die Maximierung der Korrelationen $r(\vec{U}_s, \vec{V}_s)$ ist.

Wegen (6) und (7) kann man die Korrelation $r(\vec{U}_s, \vec{V}_s)$ in der Form $r(\vec{U}_s, \vec{V}_s) = \vec{a}_s' X'Y\vec{b}_s$ schreiben. Zusammen mit der Nebenbedingung (10) hat man dann

$$r(\vec{U}_1, \vec{V}_1) = \frac{\vec{a}_1' X'Y\vec{b}_1}{\sqrt{(\vec{a}_1' X'X\vec{a}_1)(\vec{b}_1' Y'Y\vec{b}_1)}} \stackrel{!}{=} \max, \quad \vec{a}_1' X'X\vec{a}_1 = \vec{b}_1' Y'Y\vec{b}_1 = 1. \tag{11}$$

Für $\vec{U}_2, \vec{V}_2$ etc. ist die Formel analog. Die in (11) verwendete Schreibweise kann zur besseren Übersichtlichkeit vereinfacht werden. Da X und Y als standardisiert vorausgesetzt werden, stehen in der Matrix $X'Y/N$ alle Korrelationen zwischen den X- und den Y-Variablen. Deswegen schreibt man auch R_{xy} für $X'Y/N$. In $X'X/N$ bzw. $Y'Y/N$ stehen alle Korrelationen zwischen entweder den X-Variablen oder den Y-Variablen, so dass man auch R_{xx} und R_{yy} für $X'X/N$ und $Y'Y/N$ schreibt. Statt $r(U_1, V_1)$ kann man auch allgemein $R_{u_1v_1}$ schreiben. Die Gleichung (11) nimmt dann die äquivalente Form

$$R_{u_1v_1} = \frac{\vec{a}_1' R_{xy}\vec{b}_1}{\sqrt{(\vec{a}_1' R_{xx}\vec{a}_1)(\vec{b}_1' R_{yy}\vec{b}_1)}} \stackrel{!}{=} \max, \quad \vec{a}_1' R_{xx}\vec{a}_1 = \vec{b}_1' R_{yy}\vec{b}_1 = 1 \tag{12}$$

an, was wegen der Nebenbedingung natürlich äquivalent mit

$$R_{u_1v_1} = \vec{a}_1' R_{xy}\vec{b}_1 \stackrel{!}{=} \max \tag{13}$$

ist (N kürzt sich heraus!). Für $R_{u_2v_2}$ gilt die analoge Definition mit der Forderung, dass (i) $\vec{U}_2$ und $\vec{V}_2$ orthogonal zu $\vec{U}_1$ bzw. $\vec{V}_2$ sind, und (ii) $R_{u_2v_2}$ wiederum maximal relativ zu den noch verbleibenden kanonischen Korrelationen ist.

Die kanonischen Korrelationen sind Maße für die Übereinstimmung zwischen den korrespondierenden latenten Dimensionen. Ein weiteres Maß für die Übereinstimmung zwischen den beiden Variablensätzen ist die *Redundanz*; die Redundanz ist im Prinzip ein gewichteter Mittelwert der kanonischen Korrelationen. Autoren wie Rencher (1995)[4] kritisieren das Redundanzmaß allerdings als wenig ergiebig und argumentieren, dass man sich auf die kanonischen Korrelationen konzentrieren solle. Es soll an dieser Stelle deswegen nicht weiter darauf eingegangen werden.

2.2 Die Bestimmung der Gewichtungsvektoren

R_{uv} hängt von den unbekannten Vektoren $\vec{a}$ und $\vec{b}$ ab. Sie müssen so bestimmt werden, dass R_{uv} maximal wird. Dazu kann man den Ausdruck (11) für R_{uv} nach $\vec{a}$ und $\vec{b}$ differenzieren, die entstehenden Ableitungen gleich Null setzen und nach $\hat{a}$ und $\hat{b}$ auflösen, wobei $\hat{a}$ und $\hat{b}$ diejenigen Vektoren sind, für die die Ableitungen eben Null werden. Es wird zunächst das Ergebnis dieser Berechungen genannt:

Satz 1 *Die Vektoren $\vec{a}_k$ und $\vec{b}_k$, $k = 1, \ldots, \min(p, q)$ sind durch die Gleichungen*

$$R_{xx}^{-1} R_{xy} R_{yy}^{-1} R_{yx} \vec{a}_k = \lambda_k \vec{a}_k \tag{14}$$

und

$$R_{yy}^{-1} R_{yx} R_{xx}^{-1} R_{xy} \vec{b}_k = \lambda_k \vec{b}_k \tag{15}$$

gegeben, d.h. als Eigenvektoren der Matrizen

$$R_{xx}^{-1} R_{xy} R_{yy}^{-1} R_{yx} \quad \textit{und} \quad R_{yy}^{-1} R_{yx} R_{xx}^{-1} R_{xy}.$$

Die zugehörigen kanonischen Korrelationen $R(\vec{U}_k, \vec{V}_k)$ sind durch die Wurzeln der Eigenwerte λ_k gegeben, d.h. es gilt

$$R(\vec{U}_k, \vec{V}_k) = \sqrt{\lambda_k}, \quad k = 1, 2, \ldots, r = \min(p, q). \tag{16}$$

Beweis: Den Beweis findet man in Abschnitt 5.2.1. □

Nach (6) und (7) erhält man dann die Scores der Personen gemäß

$$U = XA, \quad V = YB, \tag{17}$$

wobei die verschiedenen Vektoren $\vec{a}_s$ und $\vec{b}_s$ zu entsprechenden Matrizen A und B zusammengefaßt werden. Die Spaltenvektoren von U und V haben die folgenden Eigenschaften:

Satz 2 *Es seien $\vec{U}_j$, $\vec{U}_k$, $\vec{V}_j$ und $\vec{V}_k$ kanonische Variablen, $j, k = 1, \ldots, \min(p, q)$. Dann gelten die folgenden Aussagen:*

$$\begin{aligned} \vec{U}_j{}'\vec{U}_k &= 0, \quad j \neq k && (18)\\ \vec{V}_j{}'\vec{V}_k &= 0, \quad j \neq k && (19)\\ \vec{U}_j{}'\vec{V}_k &= 0, \quad j \neq k && (20)\\ \vec{U}_j{}'\vec{V}_j &= R_j, \quad j = k && (21) \end{aligned}$$

wobei $R_j = R(\vec{U}_j, \vec{V}_j) = \sqrt{\lambda_j}$, und λ_j ist in (15) und (14) definiert worden.

[4] Rencher, A.: Methods of Multivariate Analysis. New York, John Wiley 1995

Beweis: Den Beweis findet man in Abschnitt 5.2.2. □

2.3 Zur Deutung der Gewichtsvektoren $\vec{a}_j$ und $\vec{b}_k$

Die Komponenten der Vektoren $\vec{a}_j$ und $\vec{b}_k$ sind wie Regressionsgewichte zu interpretieren, mit denen $\vec{U}_j$ bzw. $\vec{V}_k$ "vorausgesagt" werden; in Abschnitt 4 wird diese Interpretation illustriert. Da sich diese Gewichte auf latente Dimensionen beziehen, könnte man vermuten, dass die Komponenten der $\vec{a}_j$ etc wie Faktorladungen interpretiert werden können. Dies ist nicht ganz der Fall, wie im Folgenden gezeigt werden soll.

Die Komponenten der Spaltenvektoren von A und B (dh die $\vec{a}_j$ und $\vec{b}_k$) enthalten die Gewichte der Variablen $X_1, \ldots, X_p$, die zur Berechnung der (Spalten-)Vektoren von U bzw. V benötigt werden. Die Vektoren in U und V sind jeweils orthogonal. Bei der Hauptachsentransformation hat man ebenfalls die Transformation der Vektoren in X in orthogonale Vektoren L, dh $L = XT$, und aus der geforderten Orthogonalität von L folgt $L'L = \Lambda = T'X'XT$, Λ eine Diagonalmatrix, woraus wiederum folgt, dass T die orthogonale Matrix der Eigenvektoren von $X'X$ ist. Wegen der geforderten Orthogonalität der Vektoren U (und V) folgt wiederum $U'U = A'X'XA = \Lambda_u$ bzw. $V'V = B'X'XB = \Lambda_v$, wobei Λ_u und Λ_v Diagonalmatrizen sind. Man könnte versucht sein, daraus die Orthogonalität der Vektoren in A und B zu folgern. Dieser Schluß ist aber nicht korrekt, da er die Nebenbedingungen für A und B vernachlässigt. Nach (14) sind die Spaltenvektoren von A ja die Eigenvektoren der Matrix $C = R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}$. Die Eigenvektoren dieser Matrix sind orthogonal, wenn C symmetrisch ist. Ist aber C nicht symmetrisch, so kann die Orthogonalität der Spaltenvektoren von A nicht mehr gefolgert werden, dh die Vektoren können auch nicht orthogonal sein. Man macht sich diesen Sachverhalt klar, indem man im Prinzip wie bei dem Beweis der Orthogonalität der Eigenvektoren einer symmetrischen Matrix vorgeht. So müssen nach (14) für irgendzwei, also etwa $\vec{a}_1$ und $\vec{a}_2$ die Gleichungen

$$R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}\vec{a}_1 = \lambda_1\vec{a}_1 \tag{22}$$

$$R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}\vec{a}_2 = \lambda_2\vec{a}_2 \tag{23}$$

gelten. Man kann nun (22) mit $\vec{a}_2'$ und (23) mit $\vec{a}_2'$ multiplizieren:

$$\vec{a}_2\,'R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}\vec{a}_1 = \lambda_1\vec{a}_2'\vec{a}_1 \tag{24}$$

$$\vec{a}_1\,'R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}\vec{a}_2 = \lambda_2\vec{a}_1'\vec{a}_2 \tag{25}$$

Man kann nun die linke und die rechte Seite von (25) stürzen und erhält

$$\vec{a}_2'R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}\vec{a}_1 = \lambda_1\vec{a}_2'\vec{a}_1 \tag{26}$$

$$\vec{a}_2'R_{yx}'R_{yy}^{-1}R_{xy}'R_{xx}^{-1}\vec{a}_1 = \lambda_2'\vec{a}_2'\vec{a}_1 \tag{27}$$

Zur Abkürzung werde $C = R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}$ gesetzt; C ist eine $p \times p$-Matrix, also ist auch C' eine $p \times p$-Matrix, so dass die Differenz $C - C'$ wohldefiniert ist. Subtrahiert man die zweite Gleichung von der ersten, so resultiert

$$\vec{a}_2'(C - C')\vec{a}_1 = \vec{a}_2'\vec{a}_1(\lambda_1 - \lambda_2). \tag{28}$$

Es werde angenommen, dass $\vec{a}_2\,'(C - C')\vec{a}_1 = 0$ ist; dann ist auch $\vec{a}_2\,'\vec{a}_1(\lambda_1 - \lambda_2) = 0$. Setzt man $\lambda_1 \neq \lambda_2$ voraus, so muß dann $\vec{a}_2'\vec{a}_1 = 0$ gelten, dh die Vektoren $\vec{a}_1$ und $\vec{a}_2$ sind orthogonal. Die Frage ist aber, ob $\vec{a}_2{}'(C - C')\vec{a}_1 = 0$ allgemein gilt. Die Matrix C ist ja i.a. nicht symmetrisch, so dass man nicht $C - C' = 0$ annehmen kann, und damit kann man die Bedingung $\vec{a}_2{}'(C - C')\vec{a}_1 = 0$ nicht allgemein als erfüllt ansehen. Deswegen kann man auch nicht folgern, dass $\vec{a}_1{}'\vec{a}_2 = 0$ gelten muß, - die beiden Vektoren sind also nicht notwendig orthogonal! Für eine nichtsymmetrische Matrix C ist die Orthogonalität der beiden Eigenvektoren also ein eher unwahrscheinlicher Spezialfall. Die Betrachtungen für die Eigenvektoren in B sind analog.

Die Konsequenz aus diesen Betrachtungen ist, dass man die Komponenten der Vektoren $\vec{a}_1$, $\vec{a}_2$, etc nicht als "Ladungen" der entsprechenden Variablen auf orthognalen latenten Dimensionen interpretieren darf.

3 Signifikanztests

Signifikanztests können bei der Interpretation der Resultate hilfreich sein. Es sei $R_s^2 = R^2_{U_s V_s}$, und es werde

$$\Lambda = \prod_{k=1}^{\min(p,q)} (1 - R_s^2). \tag{29}$$

(Hier ist Λ anders definiert als in der Singularwertzerlegung; Λ bedeutet in diesem Abschnitt einfach die rechte Seite von (29)). Es kann gezeigt werden, dass Λ approximativ χ^2-verteilt ist:

$$\chi^2 \approx -N(-\frac{1}{2}(p + q + 3) \log \Lambda, \quad df = pq. \tag{30}$$

Alternativ dazu kann man den F-Quotienten

$$F = \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \frac{df_2}{df_1}, \quad df_1 = pq, \quad df_2 = wt - .5pq + 1 \tag{31}$$

mit

$$w = N - .5(p + q + 3), \quad t = \sqrt{\frac{(pq)^2 - 4}{p^2 + q^2 - 5}}$$

benutzen. Es sei

$$\Lambda_j = \prod_{k=j}^{\min(p,q)} (1 - R_k^2) = \prod_{k=j}^{\min(p,q)} (1 - \lambda_k), \tag{32}$$

und

$$\chi_j^2 \approx -N(-\frac{1}{2}(p + q + 3) \log \Lambda_j, \quad df = (p - j + 1)(q - j + 1). \tag{33}$$

Mit χ_j^2 lassen sich die einzelnen kanonischen Korrelationen auf Signifikanz prüfen (für den F-Test gibt es einen analogen Ausdruck, vergl. Rencher (1995), p. 404): der Test (30) liefert einen globalen Signifikanzwert. Die Frage ist dann, ob vielleicht nur der erste Korrelationskoeffizient signifikant ist. Wird χ_2^2 ebenfalls signifikant, so ist der zweite Koeffizient ebenfalls signifikant, etc.

4 Beispiele für Kanonische Korrelationen

Beispiel 1 Es werden die Daten aus der Befragung im WS 04/05 betrachtet. Man kann die Hypothese aufstellen, dass die Beurteilung des Dozenten anhand der Antworten der Studierenden auf die ersten drei Fragen ("Vorwissen", "Interesse", "beruflich wichtig") vorhergesagt werden kann, - und umgekehrt. Die Variablenmengen S_1 und S_2 werden dementsprechend in der Tabelle 1 zusammengefasst. Die Tabelle enthält auch die Kanonischen Gewichte, dh die Komponenten der Vektoren $\vec{a}_1$ und $\vec{b}_1$. Die Items in den jeweiligen Mengen sind nach der Größe der Kanonischen Gewichte geordnet; mehr als eine Dimension muß bei diesen beiden Datensätzen nicht berücksichtigt werden, denn nur die erste Kanonische Korrelation $R_1 = .837$ ist signifikant. Die Kanonischen Variablen $\vec{U}_1$ und $\vec{V}_1$ sind dementsprechend wie folgt definiert:

Tabelle 1: Variablen für die Analyse, WS 04/05

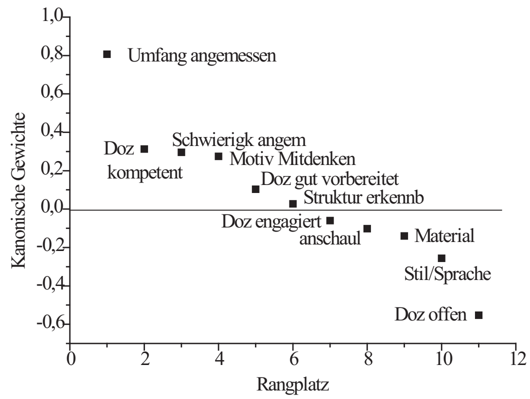
Variablen S_1		$\vec{a}$	Variablen S_2		$\vec{b}$
X_2	Interesse für d. Stoff	.479	Y_1	Umfang angemessen	.807
X_1	Vorwissen hinreichend	.407	Y_6	Doz ist kompetent	.313
X_3	Stoff ist berufl. wichtig	.397	Y_2	Schwierigk. ist angemessen	.276
			Y_{10}	Doz. motiv. z. Mitdenken	.275
			Y_5	Doz. gut vorbereitet	.103
			Y_3	Veranstalt. ist strukturiert	.027
			Y_7	Doz. ist engagiert	-.060
			Y_4	Darstell. ist anschaulich	-.102
			Y_9	Material ist gut	-.140
			Y_8	Vortragsstil/Sprache gut	-.256
			Y_{11}	Doz. offen für Fragen etc	-.552
$R_1 = .837,\ R_2 = .419,\ R_3 = .391$					

$$\vec{U}_1 = .407\vec{X}_1 + .479\vec{X}_2 + .397\vec{X}_3 \quad (34)$$

$$\vec{V}_1 = .807\vec{Y}_1 + .276\vec{Y}_2 + \cdots + .275\vec{Y}_{10} - .552\vec{Y}_{11} \quad (35)$$

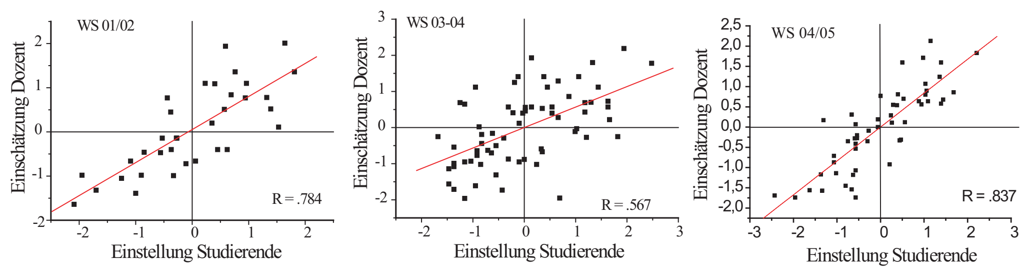
Die Kanonischen Gewichte der Menge S_1 haben ungefähr die gleiche Größenordnung, wobei das Interesse für den Stoff eine dominantere Rolle spielt als die Einschätzungen des eigenen Vorwissens und der beruflichen Wichtigkeit. Abbildung 1 gibt einen graphischen Eindruck vom Verlauf der Gewichte für die Menge S_2: die x-Achse repräsentiert die Rangordnung eines Items (1 bis 11), die y-Achse das jeweilige Kanonische Gewicht. Das mit Abstand wichtigste Item ist hier die Einschätzung der Angemessenheit des Stoffumfangs. Wer den Umfang des Stoffes als angemessen empfindet, der gibt auch hohe Schätzungen für die S_1-Items ab, und wer den Stoffumfang als unangemessen (zu hoch! - auch ein zu geringer Stoffumfang könnte ja als unangemessen gelten, aber diese Einschätzung wird hier kaum vorgekommen sein) ansieht, wird niedrige Ratings bei den S_1-Items, insbesondere für das Vorwissen abgeben (dies sind zusammenfassende statistische Aussagen, einzelne Studierende weichen von diesem Schema ab). Wer hohe Schätzungen für

Abbildung 1: Kanonische Gewichte, vergl. Tabelle 1.



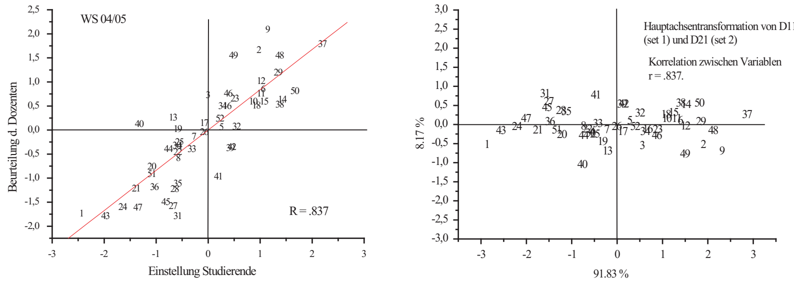
die S_1-Items abgibt, gibt auch positive Einschätzungen für die Kompetenz des Dozenten, die Angemessenheit der Schwierigkeit und für die Motivation zum Mitdenken ab. Ob der Dozent als gut vorbereitet erscheint hat nur geringes Gewicht, und bemerkenswert sind die Gewichte der Einschätzungen für die Struktur der Darstellung und das Engagement des Dozenten: diese beiden Items scheinen gar keine Rolle (für die Vorhersage der S_1-Items) zu spielen. Die Anschaulichkeit hat ebenfalls ein überraschend geringes, allerdings negatives Gewicht. Wer auf Anschaulichkeit Wert legt, signalisiert damit geringe Werte auf denS_1-Items, oder anders herum formuliert: wer niedrige Werte auf den Merkmalen "Vorwissen", "Interesse" und "beruflich wichtig" angibt, ist ganz besonders auf die Anschaulichkeit der Darstellung der Verfahren angewiesen). Dies gilt auch für restlichen S_2-Items, insbesondere für diejenigen Studierenden, die die Offenheit des Dozenten hoch einschätzen, zeigen damit niedrige Werte auf den S_1-Items an. Das ist einleuchtend: die Verständnisschwierigkeiten werden um so größer, je geringer die Vorkenntnisse sind, und damit spielt die Möglichkeit, Fragen zu stellen, eine bedeutende Rolle. Abb. 2 zeigt den

Abbildung 2: Kanonische Variablen; zum Vergleich werden auch die entsprechenden Plots aus vorangegangenen Semestern gezeigt.



Plot der Kanonischen Variablen V_1 (Einschätzung des Dozenten) versus U_1 (Selbsteinschätzung des Vorwissens etc), zusammen mit den entsprechenden Abbildungen für die vorangegangenen Semester; jeder Punkt entspricht einer Person. Für das WS 04/05 ist der Zusammenhang zwischen S_1 und S_2 mit einem $R = .837$ am besten etabliert. Es fällt auf, dass es einige Studierende gibt, die sich selbst eher mittlere Noten auf denS_1-Items geben, den Dozenten aber besonders schlecht beurteilen, und ebenso so Studierende, die sich zwar in den S_1-Items als eher überdurchschnittlich gut einschätzen, den Dozenten gleichwohl aber besonders gut beurteilen. Abbildung 3 zeigt die Daten für das WS 04/05

Abbildung 3: Plot der Kanonischen Variablen (a), Hauptachsentransformation der Kanonischen Variablen (b)



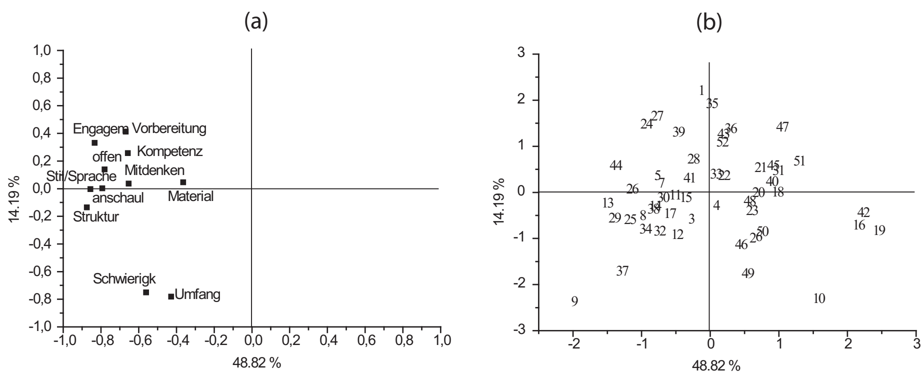
noch einmal, wobei jetzt aber die Punkte durch Zahlen ersetzt sind; diese für den dazu korrespondierenden Punkt in Abb. 2, dh für eine Person (in der Reihenfolge, in der die Personen in der Ausgangsdatenmatrix auftauchen). Die Darstellung soll helfen, die Position einer gegebenen Person in verschiedenen Räumen zu identifizieren, vergl. Abb. 4 weiter unten. In Abb. 3, (a) wird einfach noch einmal der Plot für das WS 04/05 aus Abb. 2 gezeigt. Wie bei einer Regression liegen hier die Punkte/Zahlen um die Regressionsgerade, die die Beziehung zwischen den ersten latenten Dimensionen von S_1 und S_2 abbildet. Diese Regressionsgerade wirkt wie die erste Hauptachse des Ellipsoids, das die Punkte/Zahlen bilden. Eine Hauptachsentransformation der Matrix, deren erste Spalte die Koordinaten der Personen auf der ersten Dimension für S_1 und deren zweite Spalte die Koordinaten der Personen auf der ersten Dimension für S_2 sind, sollte die Regressionsgerade in Abb. 2 im wesentlichen reproduzieren. So ist es auch, wie Abb. 3 (b) zeigt. Die erste Hauptachse erklärt 91.8% der Varianz in diesen Daten, während die zweite Hauptachse nur noch 8.2% erklärt.

Die Korrelationen R_{xx}, R_{xy}, R_{yx} und R_{yy} zwischen den X- und Y-Variablen sind in Tabelle 2 angegeben worden, wobei die Korrelationen R_{yx} zwischen den Y- und den X-Variablen nicht mehr angeschrieben wurden, da sie sich wegen $r_{yx} = r_{xy}$ (also $R_{yx} = R'_{xy}$) aus den Korrelationen R_{xy} zwischen den X- und den Y-Variablen ergeben. Dass sich der Zusammenhang zwischen den X- und den Y-Variablen durch eine latente Dimension erklären läßt ist im wesentlichen ein Resultat der Tatsache, dass nur drei X-Variablen betrachtet wurden, die sich sehr gut durch eine Dimension ausdrücken lassen; 61 % der Varianz dieser drei Variablen lassen sich durch die erste latente Variable für diese drei ausdrücken, 21.4 % und 17.4% für die restlichen beiden, wie sich in einer Hauptkomponentenanalyse, die hier nicht ausführlich vorgestellt wird, nachweisen läßt. Die Frage ist, ob diese Aussage auch für die Y-, also die Dozentenvariablen gilt. Zu diesem Zweck ist noch einmal eine Hauptkomponentenanalyse nur für die Y-Variablen gerechnet worden, deren Resultat in Abbildung 4 vorgestellt wird. Die erste Hauptachse für diese Variablen erklärt ca 49 % der Varianz dieser Variablen, und die zweite nur noch ca 14 % dieser Variablen. Der Vergleich mit Abbildung 3 zeigt, dass die Hauptachsen dieser Variablen nicht mit der Dimension, die bei der Bestimmung der Kanonischen Korrelation für diese Daten berechnet wird, übereinstimmt, wie man durch Vergleich der Positionen der die Personen bezeichnenden Zahlen leicht feststellt. Ordnet man die Variablen in S_2 in der

Tabelle 2: Korrelationen R_{xx}, R_{xy}, R_{yx} und R_{yy}

	Variablen S_1			Variablen S_2										
	X_1	X_2	X_3	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}
X_1	1.000			.575	.439	.228	.198	.013	.011	-.092	-.013	.066	-.022	-.112
X_2	.365	1.000		.531	.385	.054	.117	-.149	.082	-.167	.035	-.072	.203	-.096
X_3	.463	.411	1.000	.607	.357	.038	.015	-.076	.024	-.042	.068	.036	.000	-.142
Y_1				1.000										
Y_2				.718	1.000									
Y_3				.349	.382	1.000								
Y_4				.366	.626	.758	1.000							
Y_5				.057	.085	.673	.566	1.000						
Y_6				.156	.155	.384	.488	.392	1.000					
Y_7				.134	.235	.603	.623	.704	.525	1.000				
Y_8				.350	.451	.632	.714	.476	.519	.774	1.000			
Y_9				.085	.207	.286	.242	.170	.340	.242	.120	1.000		
Y_{10}				.190	.299	.378	.512	.237	.404	.507	.548	.284	1.000	
Y_{11}				.212	.361	.435	.592	.453	.582	.681	.647	.222	.561	1.000

Abbildung 4: Hauptkomponentenanalyse der Variablen in S_2; (a) Variablen (Skalen), (b) Personen.



Reihenfolge ihrer Projektionen auf die erste Hauptachse an, so zeigt der Vergleich der Tabelle 1 (oder Abbildung 1) mit der Tabelle 3, dass die beiden Rangordnungen nicht miteinander übereinstimmen. Man sieht weiter, dass die Faktoren- bzw. Hauptkomponentenanalyse mehr als nur eine latente Variable für die jeweiligen Datensätze nahelegen; in diesem Sinne stellt die Kanonische Korrelation *keine* gute Zusammenfassung der Daten dar. Das Ziel der Kanonischen Korrelation ist auch ein anderes: es sollen ja eben nur diejenigen latenten Variablen aus den beiden Datensätzen betrachtet werden, die maximal miteinander korrelieren. Neben der Kanonischen Korrelation wird man immer auch die faktorenanalytischen Ergebnisse für die einzelnen Datensätze diskutieren wollen.

Es sei noch einmal darauf hingewiesen, dass die Darstellungen (a) der Items und (b) der Personen in Abbildung 4 aufeinander bezogen werden können. Das Cluster der Items von Struktur bis Engagement und Vorbereitung signalisiert, dass diese Items relativ hoch miteinander korrelieren, ebenso korrelieren die beiden Items Umfang und Schwierigkeit (sind angemessen) relativ hoch. Personen, die in (b) den Positionen der Items in (a)

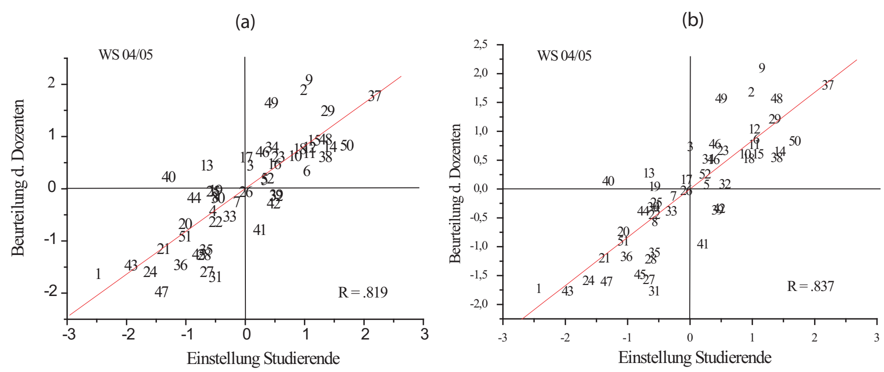
Tabelle 3: Rangfolge der Projektionen der Variablen S_2 auf die erste Hauptachse

Anschaul	Stil/Sprache	Engagement	Struktur	Offenheit	Vorbereitung
-.875	-.856	-.843	-.793	-.781	-.669
Kompetenz	Mitdenken	Schwierigkeit	Umfang	Material	
-.657	-.653	-.561	-.428	-.364	

Tabelle 4: Variablen in der Analyse der reduzierten S_2-Menge

Variablen S_1		$\vec{a}$	Variablen S_2		$\vec{b}$
X_3	wichtig	.457	Y_1	Umfang	.865
X_2	Interesse	.424	Y_6	Kompetenz	.205
X_1	Vorwissen	.399	Y_{10}	Mitdenken	.156
			Y_2	Schwierigkeit	.138
			Y_{11}	Doz. offen	-.624
$R_1 = .819, \;\; R_2 = .275, \;\; R_3 = .145$					

Abbildung 5: (a) Kanonische Korrelation für die reduzierte S_2-Menge, vergl. Tabelle 4, (b) für die vollständige S_2-Menge.



entsprechen haben i.a. hohe Ratings für diese Items abgegeben. So haben die Personen 9 und 37 hohe Ratings für die Angemessenheit des Stoffumfangs und der Schwierigkeiten angegeben (zu den anderen Items ebenfalls). Die Personen 24 und 27 haben insbesondere das Engagement und die gute Vorbereitung des Dozenten gelobt (und gute Ratings für die übrigen Items abgegeben). Die Personen 10, 16, 19 und 42 beurteilen den Dozenten auf praktisch allen Skalen besonders schlecht. Insofern kann man die Abbildungen (a) und (b) wie einen Biplot interpretieren. Eine Zusammenfassung in *einen* Graphen wird aber nicht betrachtet: man beachte, dass die Skalen der jeweiligen Koordinatenachsen verschieden sind. Für einen Biplot müßten diese Skalen kompatibel gemacht werden.

Zur Illustration ist die Kanonische Analyse mit einer reduzierten S_2-Menge noch einmal durchgeführt worden; Tabelle 4 enthält die betrachteten Variablen für die reduzierte S_2-Menge sowie deren Gewichte.

Abb. 5 (a) zeigt das Resultat für die reduzierte S_2-Menge. Es ergibt sich immer noch eine Kanonische Korrelation von $R_1 = .819$. Zum Vergleich ist das Resultat für die nicht reduzierte Menge nocheinmal mit abgebildet worden. Die Inspektion der beiden Abbildungen zeigt, dass sich die Positionen der Personen nur geringfügig für die beiden S_2-Mengen unterscheiden. Die Rangordnungen der Items hinsichtlich ihrer Gewichte in Tabelle 4 unterscheidet sich, wenn auch nur geringfügig, von der der ersten Analyse (Abb. 1). Die S_2-Variablen der Tabelle enthalten praktisch schon die Information der Menge der Variablen in Tabelle 1. □

Beispiel 2 Berufliche Zufriedenheit: Dunham (1977)[5] betrachtete den Effekt der Organisationsstruktur auf die "job satisfaction". Dabei wurden Komponenten der Zufriedenheit auf die der Job-Charakteristika bezogen. Es wurden $N = 784$ Mitarbeiter befragt. Tabelle 5 enthält die Variablen zusammen mit den geschätzen Kanonischen Gewichten.

Tabelle 5: Variablen zu Job-Charaktistiken (S_1) und zur Job-Zufriedenheit (S_2)

Variablen S_1		$\vec{a}_1$	Variablen S_2		$\vec{b}_1$
X_5	autonomy	.44	Y_6	kind-of-work-satisfaction	.52
X_1	feedback	.44	Y_1	supervisor satisfaction	.41
X_2	task significance	.21	Y_5	company identification	.29
X_3	task variety	.17	Y_2	career-future satisfaction	.22
X_4	task identity	-.02	Y_4	workload satisfaction	.01
			Y_3	financial satisfaction	-.03
			Y_7	general satisfaction	-.12

Die Korrelationen

$$R = \begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix}$$

zwischen den Variablen findet man in Tablle 6: Es kann maximal $\min(p, q) = \min(5, 7) = 5$ Kanonische Korrelationskoeffizienten geben. Für die Gewichtsvektoren $\vec{a}_k$ und $\vec{b}_k$ ergeben sich die Werte der Tabelle 7.

Die Vektoren $\vec{a}_k$ haben fünf Komponenten, da es fünf Variablen im ersten Variablensatz gibt, und die Vektoren $\vec{b}_k$ haben sieben Komponenten, weil es sieben Variablen im zweiten Variablensatz gibt. Die Kanonischen Variablen können hier nicht angegeben werden, da sie einerseits nicht veröffentlicht wurden, und weil sie jeweils $N = 784$ Komponenten haben. Für die Interpretation der Daten genügen überdies die Kanonischen Gewichte und die Kanonischen Korrelationen. Zur Illustration wird die Definition der ersten kanonischen Variablen gegeben:

$$\vec{U}_1 \quad = \quad .42\vec{X}_1 + .21\vec{X}_2 + .17\vec{X}_3 - .02\vec{X}_4 + .44\vec{X}_5 \tag{36}$$

[5] Dunham, R.B. (1977) Reaction to job characteristics: moderation effects of the organization. Academy of Management Journal, 20 (1), 42-65

Tabelle 6: Die Korrelationen zwischen den Variablen

	X_1	X_2	X_3	X_4	X_5	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
X_1	1.00	.49	.53	.49	.51	.33	.32	.20	.19	.30	.37	.21
X_2	.49	1.00	.57	.46	.53	.30	.21	.16	.08	.27	.35	.20
X_3	.53	.57	1.00	.48	.57	.31	.23	.14	.07	.24	.37	.18
X_4	.49	.46	.48	1.00	.57	.24	.22	.12	.19	.21	.29	.16
X_5	.51	.53	.57	.57	1.00	.38	.32	.17	.23	.32	.36	.27
Y_1	.33	.30	.31	.24	.38	1.00	.43	.27	.24	.34	.37	.40
Y_2	.32	.21	.23	.22	.32	.43	1.00	.33	.26	.54	.32	.58
Y_3	.20	.16	.14	.12	.17	.27	.33	1.00	.25	.46	.29	.45
Y_4	.19	.08	.07	.19	.23	.24	.26	.25	1.00	.28	.30	.27
Y_5	.30	.27	.24	.21	.32	. 34	.54	.46	.28	1.00	.35	.59
Y_6	.37	.35	.37	.29	.36	.37	.32	.29	.30	.35	1.00	.31
Y_7	.21	.20	.18	.16	.27	.40	.58	.45	.27	.59	.31	1.00

$$\vec{V}_1 = .42\vec{Y}_1 + .22\vec{Y}_2 - .03\vec{Y}_3 + .01\vec{Y}_4 + .29\vec{Y}_5 + .52\vec{Y}_6 - .12\vec{Y}_7 \tag{37}$$

Es sei darauf hingewiesen, dass $\vec{a}_1{}'\vec{a}_2 = -.196$ und $\vec{b}_1{}'\vec{b}_2 = -.284$; diese Werte illustrieren, dass die $\vec{a}_j$ und die $\vec{b}_k$ *nicht* orthogonal sind. Der Tabelle 7 entnimmt man, dass $r(U_1, V_1) = .55$, $r(U_2, V_2) = .23$, und die verbleibenden Korrelationen erscheinen vernachlässigbar. Insgesamt kann man sagen, dass das erste Paar kanonischer Variablen die meiste gemeinsame Varianz in den Daten erklärt. Zur Interpretation läßt sich wie folgt vorgehen. Der Gleichung (36) kann man entnehmen, dass $\vec{U}_1$ hauptsächlich durch $\vec{X}_1$ und

Tabelle 7: Die Komponenten der Gewichtsvektoren

A					B					Kan. Korr.	
$\vec{a}_1$	$\vec{a}_2$	$\vec{a}_3$	$\vec{a}_4$	$\vec{a}_5$	$\vec{b}_1$	$\vec{b}_2$	$\vec{b}_3$	$\vec{b}_4$	$\vec{b}_5$	$r(U,V)$	
.42	-.30	-.86	.76	.27	.42	.03	.58	.23	-.52	$r(U_1,V_1) =$	.55
.21	.65	.47	-.06	1.01	.22	-.42	-.76	.49	-.63	$r(U_2,V_2) =$	.23
.17	.85	-.19	-.12	-1.04	-.03	.08	-.41	.52	.41	$r(U_3,V_3) =$	.12
-.02	-.29	-.49	-1.14	-.16	.01	-.91	-.07	-.47	.21	$r(U_4,V_4) =$	.08
.44	-.81	.95	-.25	.32	.29	.14	.19	.34	.76	$r(U_5,V_5) =$	.05
					.52	.59	-.43	-.69	-.02		
					-.12	-.02	.92	-.37	.10		

$\vec{X}_5$ bestimmt wird, d.h. durch die Variablen "feedback" und "autonomy", vergl. Tabelle 5. Die kanonische Variable V_1 (d.h. der Vektor $\vec{V}_1$) wird hauptsächlich durch $\vec{Y}_1$: "supervisor satisfaction", $\vec{Y}_2$: "career-future satisfaction", und $\vec{Y}_6$: "Kind-of-work-satisfaction" bestimmt. Diese drei Variablen bestimmen massgeblich, was "satisfaction" bedeutet, und lassen sich mit Rückkopplung einerseits und Autonomie im Arbeitsalltag "erklären". Um eine weitere Interpretation der kanonischen Variablen zu erhalten, wurden die Korrelationen zwischen $\vec{U}_1$ und den "Prädiktoren" $\vec{X}_1, \ldots, \vec{X}_5$ einerseits und zwischen $\vec{V}_1$ und den $\vec{Y}_1, \ldots, \vec{Y}_7$ berechnet; ebenso die zwischen $\vec{U}_1$ und den $\vec{Y}_k$ und zwischen $\vec{V}_1$ und den $\vec{X}_j$. Interessant ist die Beziehung zwischen den Korrelationen zwischen den X-Variablen und $\vec{U}_1$ und einerseits und den Komponenten von $\vec{a}_1$ andererseits. Die Korrelationen sind

Tabelle 8: Korrelationen zwischen den kanonischen und den gemessenen Variablen

X-Variablen	$\vec{U}_1$	$\vec{V}_1$	Y-Variablen	$\vec{U}_1$	$\vec{V}_1$
Feedback	.83	.46	supervisor satisfaction	.42	.75
task signficance	.74	.41	career future satisfaction	.35	.65
task variety	.75	.42	financial satisfaction	.21	.39
task identity	.62	.34	workload satisfaction	.21	.37
autonomy	.85	.48	company identification	.36	.65
			kind-of-work satisfaction	.44	.80
			general satisfaction	.28	.50

von gleicher Größenordnung, während nur die erste und die fünfte Komponente von $\vec{a}_1$ diese Größenordnung hat. Von den Korrelationen her gesehen kann man sagen, dass $\vec{U}_1$ eine Art Index für die Job-Charakteristik repräsentiert. □

5 Anhang

Dieser Abschnitt liefert formale Querbezüge und Beweise, die nicht prüfungsrelevant sind. Die Beschäftigung mit diesen eher formalen Strukturen hat aber den Vorteil, begriffsvernetzend zu wirken ...

5.1 Singularwertzerlegung und kanonische Korrelation

Es seien A und B die Matrizen, deren Spaltenvektoren die $\vec{a}_s$ und $\vec{b}_s$ sind, d.h. es seien $A = [\vec{a}_1, \ldots, \vec{a}_r]$, $B = [\vec{b}_1, \ldots, \vec{b}_r]$. Die Gleichungen (14) und (15) können dann in der Form

$$R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}A = A\Lambda^2 \tag{38}$$

$$R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}B = B\Lambda^2 \tag{39}$$

geschrieben werden, wobei Λ^2 hier die Eigenwerte entält; Λ enthält also die Kanonischen Korrelationen. Die linken Seiten dieser beiden Gleichungen weisen eine gewisse Ähnlichkeit auf, so dass man vermuten kann, dass sie auf bestimmte Beziehungen zwischen A und B weisen. Es soll der folgende Satz bewiesen werden:

Satz 3 *Es sei*

$$M = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2} \tag{40}$$

und die Singularwertzerlegung von M sei

$$M = S\Delta^{1/2}T', \tag{41}$$

wobei S die orthonormalen Eigenvektoren von MM' und T die orthonormalen Eigenvektoren von $M'M$ enthält; Δ ist die Diagonalmatrix der korrespondierenden, von Null *verschiedenen Eigenwerte von MM' bzw. $M'M$. Dann sind die kanonischen Variablen A und B in (38) und (39) durch*

$$A = R_{xx}^{-1/2}S, \quad B = R_{yy}^{-1/2}T, \quad \Lambda^2 = \Delta \tag{42}$$

gegeben. Weiter gilt

$$A'R_{xx}^{1/2}A = B'R_{yy}^{1/2}B = I, \tag{43}$$

I die Einheitsmatrix.

Beweis: s. Anhang, Abschnitt 5.2.3.

Anmerkungen:

1. Die Matrix $M = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2}$ ist auf den ersten Blick für die Gleichung (38) charakteristisch; sie ist es aber ebenso für (39), denn $M' = R_{yy}^{-1/2}R_{yx}R_{xx}^{-1/2}$, und die Singularwertzerlegung von M' ergibt sich sofort als $M' = T\Delta^{1/2}S'$. Die Diagonalmatrix $\Delta^{1/2}$ enthält die kanonischen Korrelationen.

2. Über die Matrix M kann eine Beziehung zwischen den Vektoren in A und B zu den latenten Variablen von X und Y hergestellt werden. Die Matrix $M = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2}$, bzw. M', repräsentiert gewissermassen die Verknüpfung zwischen den beiden Datensätzen X und Y. Da $\Lambda = \Delta^{1/2}$ folgt, dass die kanonischen Korrelationen gerade durch die Wurzeln aus den Eigenwerten von $M'M$ bzw. MM' definiert sind.

3. Um die Struktur der Matrix M tiefer zu verstehen, kann man sie weiter analysieren. Es kann gezeigt werden (vergl. Abschnitt 5.2.4), dass

$$M = P_xQ_x'Q_yP_y'. \tag{44}$$

   gilt, wobei sich P_x und Q_x aus der Singularwertzerlegung von Z_x, und P_y und Q_y sich aus der Singularwertzerlegung von Z_y ergeben. Z_x und Z_y sind die (spalten-)standardisierten Matrizen X und Y, für die ja $Z_x = Q_x\Lambda_x^{1/2}P_x'$ und $Z_y = Q_y\Lambda_y^{1/2}P_y'$ gilt. (44) kann in zweifacher Weise gelesen werden: einmal kann der "Kern" der Gleichung in der Matrix $Q_x'Q_y$ gesehen werden; diese Matrix enthält die Korrelationen zwischen den "latenten" Variablen von X und den "latenten" Variablen von Y. Diese Korrelationen werden aber noch durch Prämultiplikation mit P_x und Postmultiplikation mit P_y transformiert. Andererseits kann man auch von den transformierten latenten Variablen von X und Y ausgehen. Tatsächlich kann ja Q_xP_x' als eine Rotation der Vektoren in Q_x aufgefasst werden, und analog dazu ist Q_yP_y' eine Rotation der Vektoren in Q_y. Die Matrix M enthält dann die Korrelationen zwischen diesen transformierten Vektoren, d.h. $M = (P_xQ_x')(Q_yP_y')$. Die Transformation mit P_x' bzw. P_y' hat also gerade die Eigenschaft, zu den maximal korrelierenden latenten Vektoren von X und Y zu führen!

4. Sind die X-Variablen korreliert und ebenso die Y-Variablen, so gilt $R_{xx} \neq I$, $R_{yy} \neq I$, und die Beziehung (43) bedeutet, dass die Vektoren in A und B *nicht* orthonormal sind, denn wären sie orthonormal, würde $A'A = B'B = I$ gelten. Die Orthogonalität von A oder B ist aber als Spezialfall möglich: Es sei $R_{xx} \approx I$, I

wieder die Einheitsmatrix. Dann sind die Variablen aus dem X-Variablensatz untereinander unkorreliert und (42) impliziert, dass $A \approx S$, d.h. in diesem Fall werden die Vektoren $\vec{a}_j$ in guter Näherung orthogonal sein, da ja die Matrixc S orthogonal ist. Eine analoge Aussage gilt für die Vektoren $\vec{b}_k$ ind B. Die Gewichte, d.h. die Komponenten der $\vec{a}_j$ und $\vec{b}_k$, werden dann nur durch die Korrelationen zwischen den X- und Y-Variablen bestimmt.

Die Beziehung (41) erlaubt es, sich die Bedeutung kanonischer Korrelationen noch einmal zu verdeutlichen. Dazu werde Spezialfall $R_{xx} = I$ und $R_{yy} = I$ betrachtet, d.h. es werde angenommmen, dass die X- und die Y-Variablen unkorreliert seien. Dann gilt auch $R_{xx}^{-1/2} = I$ und $R_{yy}^{-1/2} = I$ und es folgt

$$M = R_{xy} = S\Delta T',$$

und weiter

$$A = S, \quad B = T,$$

d.h. die Gewichtsvektoren $\vec{a}_k$ und $\vec{b}_k$ sind durch die Eigenvektoren S von $R_{xy}R'_{xy} = R_{xy}R_{yx}$ bzw. durch die Eigenvektoren T von $R'_{xy}R_{xy} = R_{yx}R_{xy}$ gegeben. Andererseits gilt nach (44) $M = P_xQ'_xQ_yP'_y$, wobei P_x und P_y die Eigenvektoren von R_{xx} bzw. R_{yy} sind. Gilt aber $R_{xx} = I$ und $R_{yy} = I$, so sind P_x und P_y selbst Einheitsmatrizen, wie man leicht nachrechnet, und die zugehörigen Eigenwerte sind alle gleich 1. Also folgt

$$M = Q'_xQ_y, \quad R_{xx} = I, \; R_{yy} = I,$$

und für die SVD folgt

$$M = S\Delta^{1/2}T' = ST' = Q'_xQ_y,$$

also

$$S = Q'_x, \quad T = Q'_y.$$

Also kann man

$$U = XA' = XQ_x, \quad V = YQ_y$$

folgern, d.h. die kanonischen Variablen gehen aus den Beobachtungen X und Y durch Transformation mit den normierten latenten Variablen von X bzw. Y hervor.

Diese Interpretation läßt sich auch für den Fall aufrecht erhalten, in dem die speziellen Annahme unkorrelierter X- und Y-Variablen nicht gelten. $R_{xx}^{-1/2}$ und $R_{yy}^{-1/2}$ können dann als Gewichtung der $\tilde{S}$ und $\tilde{T}$ aufgefasst werden.

Die Beziehung zwischen A und B: Es lassen sich noch Beziehungen zwischen A und B herleiten. Aus (42) folgt nach Multiplikation von links mit $R_{yy}^{-1/2}R_{yx}$ unter Berücksichtigung von (41)

$$R_{yy}^{-1/2}R_{yx}A = R_{yy}^{-1/2}R_{yx}R_{xx}^{-1/2}S = T\Lambda S'S = T\Lambda.$$

Multipliziert man von links noch einmal mit $R_{yy}^{-1/2}$, so erhält man wegen $R_{yy}^{-1/2}T = B$ die Beziehung $R_{yy}^{-1}R_{yx}A = B\Lambda$. Analog erhält man

$$R_{xx}^{-1/2}R_{xy}B = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2}T = SV\Lambda T'T = S\Lambda,$$

und nach Multiplikation von links mit $R_{xx}^{-1/2}$ findet man $R_{xx}^{-1}R_{xy}B = A\Lambda$; zusammenfassend bekommt man

$$\begin{aligned} R_{yy}^{-1}R_{yx}A &= B\Lambda, & (45)\\ R_{xx}^{-1}R_{xy}B &= A\Lambda. & (46)\end{aligned}$$

Man beachte, dass $R_{yy}^{-1}R_{yx}$ eine $(q \times p)$-Matrix ist, und die Spaltenvektoren von A sind p-dimensional (es gibt p "Tests" im Datensatz X). Die Spaltenvektoren von B sind q-dimensional, da es q "Tests" in Y gibt. Für die ersten Spaltenvektoren $\vec{a}_1$ und $\vec{b}_1$ erhält man aus (45) insbesondere

$$R_{yy}^{-1}R_{yx}\vec{a}_1 = \lambda_1\vec{b}_1, \tag{47}$$

oder

$$\frac{1}{\lambda_1}R_{yy}^{-1}R_{yx}\vec{a}_1 = \vec{b}_1. \tag{48}$$

Man kann diese Gleichungen als Möglichkeit, die Tests in beiden Datensätzen *in einem* Koordinatensystem darzustellen, auffassen (Biplot), denn nach (42) ergeben sich die Vektoren in A und B ja durch Transformation aus den Vektoren V und W, die sich aus der SVD der Matrix M ergeben, die sich wiederum aus den Korrelationen R_{xx}, R_{yy} und R_{xy} ergibt. Die zunächst als verschieden angesehenen Koordinatensysteme A und B werden durch den Bezug auf M gleichzeitig auf ein gemeinsames Koordinatensystem definiert. (48) zeigt, wie die Konfiguration der Tests der X-Messungen relativ zu den Y-Messungen in bezug auf die erste "gemeinsame" Achse definiert ist. Es sei $C = (1/\lambda_1)R_{yy}^{-1}R_{yx}$; C ist eine $(q \times p)$-Matrix. Dann ist die i-te Komponente von $\vec{b}_1$ durch

$$b_{i1} = c_{i1}a_{11} + c_{i2}a_{21} + \cdots + c_{ip}a_{p1}. \tag{49}$$

Die Koeffizienten a_{j1}, d.h. die Komponenten von $\vec{a}_1$, sind für alle Komponenten b_{i1} die gleichen. Die Elemente der Matrix C ergeben sich gemäß

$$c_{ik} = \frac{1}{\lambda_1}\sum_{j=1}^{q} s_{ij}r_{jk}, \tag{50}$$

wobei s_{ij} das (i,j)-te Element von R_{yy}^{-1} und r_{jk} das (j,k)-te Element von R_{yx} ist; r_{jk} ist die Korrelation zwischen der Variablen Y_j und der Variablen X_j. Wegen des komplexen Aufbaus der Elemente c_{ik} ist die Beziehung schwierig in bezug auf die Korrelationen zwischen den X- und den Y-Variablen zu interpretieren. Die Beziehung spielt jedoch eine Rolle, wenn Kontingenztabellen mithilfe der Korrespondenzanalyse analysiert werden.

5.2 Beweise

5.2.1 Beweis für Satz 1

Man sieht leicht, dass dabei die Randbedingung $0 \leq R_{uv} \leq 1$ nicht notwendig berücksichtigt wird. Dies wiederum liegt daran, dass die Randbedingungen

$$\vec{u}\,'\vec{u} = \vec{a}\,'R_{xx}\vec{a} = 1, \quad \vec{v}\,'\vec{v} = \vec{b}\,'R_{yy}\vec{b} = 1 \tag{51}$$

nicht berücksichtigt worden sind. Man wird also die Ableitungen nach $\vec{a}$ und $\vec{b}$ unter den Nebenbedingungen $\vec{u}\,'\vec{u} = \vec{v}\,'\vec{v} = 1$ bilden. Dazu setzt man die Funktion

$$\begin{aligned} Q(\vec{u},\vec{v}) &= R_{uv} - \frac{1}{2}\lambda(\vec{u}\,'\vec{v} - 1) - \frac{1}{2}\mu(\vec{v}\,'\vec{v} - 1) \\ &= \vec{a}\,'\vec{x}'\vec{y}\,\vec{b} - \frac{1}{2}\lambda(\vec{a}\,'\vec{x}'\vec{x}\,\vec{a} - 1) - \frac{1}{2}\mu(\vec{b}\,'\vec{y}\,'\vec{y}\,\vec{b} - 1) \end{aligned} \tag{52}$$

an; λ und μ sind die Lagrange-Multiplikatoren; der Faktor $1/2$ wurde nur eingeführt, um die folgenden Ausdrücke zu vereinfachen (man müßte sonst mit $\lambda/2$ und $\mu/2$ rechnen!). Differenziert man Q nun nach $\vec{a}$ und $\vec{b}$ und setzt die Ableitungen gleich Null, so erhält man mit $R_{xy} = \vec{x}\,'\vec{y}$, $R_{xx} = \vec{x}\,'\vec{x}$ etc.

$$\frac{\partial Q}{\partial \vec{a}} = R_{xy}\vec{b} - \lambda R_{xx}\vec{a}|_{\vec{a}=\hat{a},\vec{b}=\hat{b}} = 0 \tag{53}$$

$$\frac{\partial Q}{\partial \vec{b}} = R_{yx}\vec{a} - \mu R_{yy}\vec{b}|_{\vec{a}=\hat{a},\vec{b}=\hat{b}} = 0. \tag{54}$$

Man bemerke, dass wegen $(\vec{x}'\vec{y})' = \vec{y}'\vec{x}$ die Beziehung $R_{yx} = R'_{xy}$ gilt. Der Übersichtlichkeit wegen soll im folgenden einfach $\vec{a}$ und $\vec{b}$ statt $\hat{a}$ und $\hat{b}$ geschrieben werden; $\vec{a}$ und $\vec{b}$ sind also ab jetzt die Vektoren, die die Bedingungen (53) und (54) erfüllen. Man kann nun die Gleichung (53) mit $\vec{a}\,'$ und die Gleichung (54) mit $\vec{b}'$ jeweils von links multiplizieren und erhält dann die Gleichungen

$$\vec{a}'R_{xy}\vec{b} - \lambda\vec{a}'R_{xx}\vec{a} = 0 \tag{55}$$

$$\vec{b}'R_{yx}\vec{a} - \mu\vec{b}'R_{yy}\vec{b} = 0 \tag{56}$$

Es ist aber $R_{uv} = \vec{a}'R_{xy}\vec{b} = \vec{b}'R_{yx}\vec{a}$ und, wegen (51), $\vec{a}'R_{xx}\vec{a} = \vec{b}'R_{yy}\vec{b} = 1$, so dass sofort

$$\lambda = \mu = R_{uv} \tag{57}$$

folgt. Die Lagrange-Multiplikatoren sind also gerade gleich der gesuchten kanonischen Korrelation. Um diese tatsächlich berechnen zu können, natürlich noch die Vektoren $\vec{a}$ und $\vec{b}$ bestimmt werden. Man kann die Gleichungen (53) und (54) unter Berücksichtigung von (57) (d.h. $\lambda = \mu$) auch in der Form

$$-\lambda R_{xx}\,\vec{a} + R_{xy}\,\vec{b} = 0 \tag{58}$$

$$R_{yx}\,\vec{a} - \lambda R_{yy}\,\vec{b} = 0 \tag{59}$$

schreiben. Man erhält eine Gleichung, die nur $\vec{a}$ enthält, wenn man zB in (58) den Vektor $\vec{b}$ eliminiert. Dazu multipliziert man (59) mit R_{yy}^{-1}:

$$R_{yy}^{-1}R_{yx}\,\vec{a} - \lambda R_{yy}^{-1}R_{yy}\,\vec{b} = R_{yy}^{-1}R_{yx}\,\vec{a} - \lambda\vec{b} = 0,$$

woraus

$$\vec{b} = \frac{1}{\lambda}R_{yy}^{-1}R_{yx}\,\vec{a}$$

folgt. Substituiert man diesen Ausdruck für $\vec{b}$ in Gleichung (58), so erhält man die Beziehung

$$-\lambda R_{xx}\,\vec{a} + R_{xy}\,\frac{1}{\lambda}R_{yy}^{-1}R_{yx}\,\vec{a} = 0,$$

bzw. $\lambda^2 R_{xx}\,\vec{a} = R_{xy}\,R_{yy}^{-1}R_{yx}\,\vec{a}$, so dass nach Multiplikation mit R_{xx}^{-1} von links

$$\lambda^2\vec{a} = R_{xx}^{-1}R_{xy}\,R_{yy}^{-1}R_{yx}\,\vec{a} \tag{60}$$

folgt; $\vec{a}$ ist also ein Eigenvektor der Matrix $R_{xx}^{-1}R_{xy}\,R_{yy}^{-1}R_{yx}$ und λ^2 ist der zugehörige Eigenwert. (60) ist aber gerade die Gleichung (14) mit $\lambda_1 = \lambda^2 = R_{uv}^2$, so dass $R_{uv} = \sqrt{\lambda_1}$ folgt. Der Beweis für (15) ist analog. □

5.2.2 Beweis für Satz 2

Man schreibt die Gleichungen (58) und (59) insbesondere für den j-ten Eigenwert und die zugehörigen Eigenvektoren $\vec{a}_j$ und $\vec{b}_j$ an,

$$\begin{aligned} -\lambda_j R_{xx}\,\vec{a}_j + R_{xy}\,\vec{b}_j &= 0 \qquad (61)\\ R_{yx}\,\vec{a}_j - \lambda R_{yy}\,\vec{b}_j &= 0 \qquad (62) \end{aligned}$$

und multipliziert dann (61) von links mit $\vec{b}'_k$, und (62) von links mit $\vec{a}'_k$, mit natürlich $j \neq k$. Es ergeben sich die Gleichungen

$$\begin{aligned} \lambda_j\vec{a}'_k R_{xx}\,\vec{a}_j + \vec{a}'_k R_{xy}\,\vec{b}_j &= 0 \qquad (63)\\ \vec{b}'_k R_{yx}\,\vec{a}_j - \lambda_j\vec{b}'_k R_{yy}\,\vec{b}_j &= 0 \qquad (64) \end{aligned}$$

Nun ist $R_{xx} = X'X$, $R_{xy} = X'Y$, etc, so dass $\vec{a}'_k R_{xx}\,\vec{a}_j = \vec{a}'_k X'X\,\vec{a}_j = \vec{U}'_k\vec{U}_j$, $\vec{a}'_k R_{xy}\,\vec{b}_j = \vec{a}'_k X'Y\,\vec{b}_j = \vec{U}'_k\vec{U}_j$, etc, so dass man die Gleichungen

$$\begin{aligned} \lambda_j\vec{U}'_k\vec{U}_j + \vec{U}'_k\vec{V}_j &= 0 \qquad (65)\\ \vec{V}'_j\vec{U}_k - \lambda_j\vec{V}'_k\vec{V}_j &= 0 \qquad (66) \end{aligned}$$

erhält; auf diese Weise erhält man ein lineares Gleichungssystem mit den unbekannten Skalarprodukten $\vec{U}'_k\vec{U}_j$, $\vec{U}'_k\vec{U}_j$ etc. Es hat drei unbekannte Skalarprodukte, also kann es noch nicht gelöst werden. Man kann nun die Gleichungen (58) und (59) nun noch einmal für $\lambda_k \neq \lambda_j$ anschreiben und genau so verfahren, wodurch man das Gleichungssystem

$$\begin{aligned} \lambda_k\vec{U}'_j\vec{U}_k + \vec{U}'_j\vec{V}_k &= 0 \qquad (67)\\ \vec{V}'_k\vec{U}_j - \lambda_k\vec{V}'_j\vec{V}_k &= 0 \qquad (68) \end{aligned}$$

erhält. Hier treten die gleichen Skalarprodukte wie in (65) und (66) auf, und man kann nun die vier Gleichungen zu einem Gleichungssystem mit vier Unbekannten anschreiben; in Matrixform ergibt sich:

$$\begin{pmatrix} \lambda_j & 1 & 0 & 0 \\ 0 & 0 & 1 & -\lambda_j \\ -\lambda_k & 0 & 1 & 0 \\ 0 & 0 & 1 & -\lambda_k \end{pmatrix} \begin{pmatrix} \vec{U}'_k\vec{U}_j \\ \vec{U}_k\vec{V}_j \\ \vec{U}_j\vec{V}_k \\ \vec{V}'_j\vec{V}_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{69}$$

Die Matrix hat aber vollen Rang, so dass es nur eine Lösung gibt, nämlich den Nullvektor, dh alle Skalarprodukte sind gleich Null. □

5.2.3 Beweis für Satz 3

R_{xx}^{-1} lässt sich nun der Form $R_{xx}^{-1/2}R_{xx}^{-1/2}$ schreiben; analog dazu hat man $R_{yy}^{-1} = R_{yy}^{-1/2}R_{yy}^{-1/2}$. Man erhält aus (38) und (39)

$$\begin{aligned} R_{xx}^{-1/2}R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2}R_{yy}^{-1/2}R_{yx}A &= A\Lambda^2 \quad (70) \\ R_{yy}^{-1/2}R_{yy}^{-1/2}R_{yx}R_{xx}^{-1/2}R_{xx}^{-1/2}R_{xy}B &= B\Lambda^2 \quad (71) \end{aligned}$$

Die angedeutete Symmetrie der linken Seiten dieser beiden Gleichungen läßt sich vervollständigen, wenn man die Matrizen

$$S = R_{xx}^{1/2}A, \quad T = R_{yy}^{1/2}B \quad (72)$$

bildet; multipliziert man von links mit $R_{xx}^{-1/2}$ bzw. $R_{yy}^{-1/2}$, so erhält man jedenfalls

$$A = R_{xx}^{-1/2}S, \quad B = R_{yy}^{-1/2}T \quad (73)$$

schreiben und diese Ausdrücke für A und B in die Gleichungen (70) und (70) einsetzen; man erhält

$$\begin{aligned} R_{xx}^{-1/2}MM'S &= R_{xx}^{-1/2}S\Lambda^2. \quad (74) \\ R_{yy}^{-1/2}M'MT &= R_{yy}^{-1/2}T\Lambda^2 \quad (75) \end{aligned}$$

Definiert man nun die Matrix M gemäß

$$M = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2}, \quad (76)$$

so folgt

$$M' = R_{yy}^{-1/2}R_{xy}R_{xx}^{-1/2}, \quad (77)$$

und man erhält wegen $R'_{yx} = R_{xy}$

$$\begin{aligned} R_{xx}^{-1/2}MM'S &= R_{xx}^{-1/2}S\Lambda^2 \quad (78) \\ R_{yy}^{-1/2}M'MT &= R_{yy}^{-1/2}T\Lambda^2, \quad (79) \end{aligned}$$

und nach Multiplikation der Gleichungen von links mit $R_{xx}^{1/2}$ bzw. mit $R_{yy}^{1/2}$

$$\begin{aligned} MM'S &= S\Lambda^2 \quad (80) \\ M'MT &= T\Lambda^2. \quad (81) \end{aligned}$$

Die SVD der Matrix M sei

$$M = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2} = S_0\Delta^{1/2}T'_0, \quad (82)$$

wobei wieder S_0 die Eigenvektoren von MM', T_0 die Eigenvektoren von $M'M$ und Δ die von Null verschiedenen Eigenwerte von MM' bzw. $M'M$ enthält. Die Gleichungen (80) und (81) implizieren $T_0 = S$, $T_0 = T$ und $\Delta = \Lambda^2$, also muss

$$M = R_{xx}^{-1/2}R_{xy}R_{yy}^{-1/2} = S\Lambda T', \quad (83)$$

gelten. (43) ergibt sich sofort aus (42) und der Tatsache, das S_0 und T_0 orthonormal sind, also $S_0 = R_{xx}^{1/2}A$ und $S'_0S_0 = I = A'R_{xx}^{1/2}R_{xx}^{1/2}A = A'R_{xx}A$, und analog für B.

5.2.4 Beweis von (44):

Es ist ja

$$R_{xx} = Z_x' Z_x = P_x \Lambda_x P_x',$$

wobei P_x die Matrix der orthonormalen Eigenvektoren von R_{xx} ist und Λ_x die Diagonalmatrix der zugehörigen Eigenwerte. Dann ist

$$R_{xx}^{-1} = (P_x \Lambda_x P_x')^{-1} = P_x \Lambda_x^{-1} P_x',$$

und

$$R_{xx}^{-1/2} = P_x \Lambda_x^{-1/2} P_x'.$$

Auf analoge Weise findet man

$$R_{yy}^{-1/2} = P_y \Lambda_y^{-1/2} P_y.$$

Weiter ist $R_{xy} = Z_x' Z_y$, und für Z_x und Z_y kann man die entsprechenden Singularwertzerlegungen einsetzen:

$$Z_x = Q_x \Lambda_x^{1/2} P_x', \quad Z_y = Q_y \Lambda_y^{1/2} P_y',$$

so dass

$$Z_x' Z_y = P_x \Lambda_x^{1/2} Q_x' Q_y \Lambda_y^{1/2} P_y'.$$

Dann findet man

$$M = P_x \Lambda_x^{-1/2} P_x' P_x \Lambda_x^{1/2} Q_x' Q_y \Lambda_y^{1/2} P_y' P_y \Lambda_y^{-1/2} P_y'.$$

Da nun

$$P_x' P_x = \Lambda_x^{-1/2} \Lambda_x^{1/2} = I, \quad P_y' P_y = \Lambda_y^{-1/2} \Lambda_y^{1/2} = I$$

ergibt sich

$$M = P_x Q_x' Q_y P_y'. \tag{84}$$

Hierin sind Q_x und Q_y die Hauptachsen für die X-Daten, und Q_y sind die Hauptachsen für die Y-Daten; sie repräsentieren diejenigen latenten Variablen für X und Y, die jeweils maximale Varianzanteile "erklären". Da Q_x und Q_y auf die Länge 1 normiert sind, kann die Matrix $Q_x' Q_y$ als Matrix der Korrelationen zwischen den latenten Variablen für X bzw. Y aufgefasst werden. $\square$

5.3 Inverse und Potenzen symmetrischer Matrizen

Um diese Beziehung herauszuarbeiten werde der Begriff der Wurzel einer symmetrischen Matrix eingeführt. Es sei M eine symmetrische, reelle $(n \times n)$-Matrix; dann kann sie durch ihre Eigenvektoren P und die zugehörigen Eigenwerte Λ dargestellt werden, d..h es gilt stets

$$M = P \Lambda P', \tag{85}$$

wobei $PP' = P'P = I$, d.h. die Eigenvektoren sind orthonormal. Man betrachte zunächst das Quadrat M^2 von M: es gilt

$$M^2 = MM = P \Lambda P' P \Lambda P' = P \Lambda^2 = P \Lambda^2 P'. \tag{86}$$

Analog dazu findet man $M^3 = P\Lambda^3 P$, und allgemein

$$M^p = P\Lambda^p P', \quad p = 1, 2, \ldots \tag{87}$$

In ähnlicher Weise kann man die Inverse von M bestimmen, vorausgesetzt, sie existiert, d.h. vorausgesetzt M hat vollen Rang. Man erinnere sich, dass wegen der Orthonormalität von P die Gleichung $P^{-1} = P'$ gilt, denn $P^{-1}P0P'P = I$. Also hat man

$$M^{-1} = (P\Lambda P')^{-1} = (P')^{-1}\Lambda^{-1}P^{-1} = P\Lambda^{-1}P'. \tag{88}$$

Man muß also nur die Inverse $\Lambda^{-1} = \mathrm{diag}(1/\lambda_1, \ldots, 1/\lambda_n)$ bilden, um M^{-1} zu berechnen. Allgemein gilt dann

$$M^{-p} = P\Lambda^{-p}P'. \tag{89}$$

Nun sei N eine symmetrische Matrix derart, dass $NN = N^2 = M$; N kann die Wurzel aus M genannt werden, so dass man $N = M^{1/2}$ schreiben kann, denn dann gilt ja $M^{1/2}M^{1/2} = NN = M$. Offenbar gilt

$$N = M^{1/2} = P\Lambda^{1/2}P', \tag{90}$$

denn

$$NN = P\Lambda^{1/2}P'P\Lambda^{1/2}P' = P\Lambda P'.$$

Analog dazu findet man

$$M^{-1/2} = P\Lambda^{-1/2}P', \tag{91}$$