Multivariate Datenanalyse

MSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(7) Kanonische Korrelationsanalyse

Modul A1/A3 Forschungsmethoden: Multivariate Verfahren | Themen

Datum	Einheit	Thema
14.10.2022	Grundlagen	(1) Einführung
21.10.2022	Grundlagen	(2) Vektoren
28.10.2022	Grundlagen	(3) Matrizen
04.11.2022	Grundlagen	(4) Eigenanalyse
11.11.2022	Grundlagen	(5) Multivariate Wahrscheinlichkeitstheorie
18.11.2022	Grundlagen	(6) Multivariate Normalverteilungen
25.11.2022	Frequentistische Inferenz	(7) Kanonische Korrelationsanalyse
02.12.2022	Frequentistische Inferenz	(8) T^2-Tests
09.12.2022	Frequentistische Inferenz	(9) Einfaktorielle MANOVA
16.12.2022	Latente Variablenmodelle	(10) Hauptkomponentenanalyse
	Weihnachtspause	
13.01.2023	Latente Variablenmodelle	(12) Lineare Normalverteilungsmodelle
20.01.2023	Latente Variablenmodelle	(13) Konfirmatorische Faktorenanalyse
27.01.2023	Latente Variablenmodelle	(14) Exploratorische Faktorenanalyse

Datenanalyseszenarien

UV	AV	Datenanalysemethoden
Univariat	Univariat	Korrelation, Einfache Regression, T-Tests
Multivariat	Univariat	Multiple Korrelation, Multiple Regression, Allgemeines Lineares Modell
Univariat	Multivariat	T^2-Tests, Einfaktorielle MANOVA
Multivariat	Multivariat	Kanonische Korrelation, Multivariates Allgemeines Lineares Modell

Datenanalyseszenarien

UV	AV
x_1	y_1
x_{11}	y_{11}
x_{12}	y_{12}
x_{13}	y_{13}
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
x_{1n}	y_{1n}

Korrelation
Einfache Regression
T-Tests

UV			AV
x_1	$\cdots$	x_m	y_1
x_{11}	$\cdots$	x_{m1}	y_{11}
x_{12}	$\cdots$	x_{m2}	y_{12}
x_{13}	$\cdots$	x_{m3}	y_{13}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	$\cdots$	x_{mn}	y_{1n}

Multiple Korrelation
Multiple Regression
Allgemeines Lineares Modell

Datenanalyseszenarien

UV	AV		
x_1	y_1	$\cdots$	y_m
x_{11}	y_{12}	$\cdots$	y_{m1}
x_{12}	y_{13}	$\cdots$	y_{m2}
x_{13}	y_{14}	$\cdots$	y_{m3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	y_{1n}	$\cdots$	y_{mn}

T^2-Tests

Einfaktorielle MANOVA

UV			AV		
x_1	$\cdots$	x_{m_x}	y_1	$\cdots$	y_{m_y}
x_{11}	$\cdots$	$x_{m_x 1}$	y_{11}	$\cdots$	$y_{m_y 1}$
x_{12}	$\cdots$	$x_{m_x 2}$	y_{12}	$\cdots$	$y_{m_y 2}$
x_{13}	$\cdots$	$x_{m_x 3}$	y_{13}	$\cdots$	$y_{m_y 3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	$\cdots$	$x_{m_x n}$	y_{1n}	$\cdots$	$y_{m_y n}$

Kanonische Korrelationsanalyse

Multivariates Allgemeines Lineares Modell

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Anwendungsszenario

Psychotherapie

Mehr Therapiestunden

$\Rightarrow$ Höhere Wirksamkeit?

Unabhängige Variable

- Anzahl Therapiestunden

Abhängige Variable

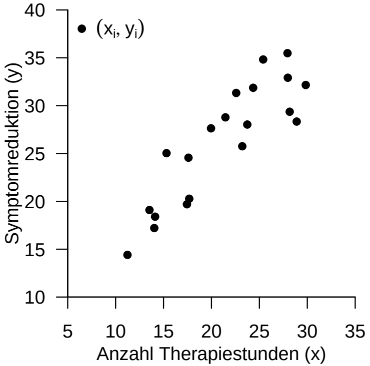
- BDI Score Reduktion

Beispieldatensatz

$i = 1, ..., 20$ Patient:innen, y_i Symptomreduktion bei Patient:in i, x_i Anzahl Therapiestunden von Patient:in i

x_i	y_i
27.9	35.5
15.3	25.0
17.4	19.7
21.5	28.8
28.2	29.4
14.0	17.2
28.0	32.9
28.9	28.3
23.2	25.8
22.6	31.3
11.2	14.4
14.1	18.4
13.5	19.1
23.7	28.0
17.7	20.3
25.4	34.8
20.0	27.6
24.4	31.9
29.8	32.2
17.6	24.6

Beispieldatensatz



Wie stark hängen Anzahl Therapiestunden und Symptomreduktion zusammen?

Definition (Korrelation)

Die *Korrelation* zweier Zufallsvariablen ξ und υ ist definiert als

$$\rho(\xi, \upsilon) := \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(\upsilon)}} \tag{1}$$

wobei $\mathbb{C}(\xi, \upsilon)$ die Kovarianz von ξ und υ und $\mathbb{V}(\xi)$ und $\mathbb{V}(\upsilon)$ die Varianzen von ξ und υ bezeichnen.

Für eine Einführung zur Korrelation siehe die entsprechenden BSc Lehreinheiten

- Erwartungswert, Varianz, Kovarianz
- Korrelation

Bemerkungen

- $\rho(\xi, \upsilon)$ wird auch *Korrelationskoeffizient* von ξ und υ genannt.
- Wir haben bereits gesehen, dass $-1 \leq \rho(\xi, \upsilon) \leq 1$ gilt.
- Wenn $\rho(\xi, \upsilon) = 0$ ist, werden ξ und υ *unkorreliert* genannt.
- Aus der Unabhängigkeit von ξ und υ folgt $\rho(\xi, \upsilon) = 0$.
- Aus $\rho(\xi, \upsilon) = 0$ folgt die Unabhängigkeit von ξ und υ im Allgemeinen nicht.

Definition (Stichprobenkorrelation)

$(x_1, y_1), ..., (x_n, y_n)$ seien Realisierungen von zweidimensionalen Zufallsvektoren $(\xi_1, \upsilon_1), ..., (\xi_n, \upsilon_n)$. Weiterhin seien:

- Die Stichprobenmittel der x_i und y_i definiert als

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \text{ und } \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{2}$$

- Die Stichprobenstandardabweichungen x_i und y_i definiert als

$$s_x := \sqrt{\frac{1}{n-1}(x_i - \bar{x})^2} \text{ und } s_y := \sqrt{\frac{1}{n-1}(y_i - \bar{y})^2}. \tag{3}$$

- Die Stichprobenkovarianz der $(x_1, y_1), ..., (x_n, y_n)$ definiert als

$$c_{xy} := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}). \tag{4}$$

Dann ist die *Stichprobenkorrelation* der $(x_1, y_1), ..., (x_n, y_n)$ definiert als

$$r_{xy} := \frac{c_{xy}}{s_x s_y} \tag{5}$$

und wird auch *Stichprobenkorrelationskoeffizient* genannt.

Beispiel

```r
# Laden des Beispieldatensatzes
fname = file.path(getwd(), "7_Kanonische_Korrelationsanalyse.csv")        # Dateipfad
D     = read.table(fname, sep = ",", header = TRUE)                       # Laden als Dataframe
x_i   = D$x_1i                                                             # x_i Werte
y_i   = D$y_1i                                                             # y_i Werte
n     = length(x_i)                                                        # n

# "Manuelle" Berechnung der Stichprobenkorrelation
x_bar = (1/n)*sum(x_i)                                                     # \bar{x}
y_bar = (1/n)*sum(y_i)                                                     # \bar{y}
s_x   = sqrt(1/(n-1)*sum((x_i - x_bar)^2))                                 # s_x
s_y   = sqrt(1/(n-1)*sum((y_i - y_bar)^2))                                 # s_y
c_xy  = 1/(n-1) * sum((x_i - x_bar) * (y_i - y_bar))                       # c_{xy}
r_xy  = c_xy/(s_x * s_y)                                                   # r_{xy}
print(r_xy)                                                                # Ausgabe
```

```
> [1] 0.883
```

```r
# Automatische Berechnung mit cor()
r_xy  = cor(x_i,y_i)                                                       # r_{xy}
print(r_xy)                                                                # Ausgabe
```

```
> [1] 0.883
```

$\Rightarrow$ Anzahl Therapiestunden und Symptomreduktion sind hochkorreliert.

Korrelation

Mechanik der Kovariationsterme

$(x_i - \bar{x}) < 0, (y_i - \bar{y}) > 0$

$\Rightarrow (x_i - \bar{x})(y_i - \bar{y}) < 0$

$(x_i - \bar{x}) > 0, (y_i - \bar{y}) > 0$

$\Rightarrow (x_i - \bar{x})(y_i - \bar{y}) > 0$

$(\bar{x}, \bar{y})$

$(x_i - \bar{x}) < 0, (y_i - \bar{y}) < 0$

$\Rightarrow (x_i - \bar{x})(y_i - \bar{y}) > 0$

$(x_i - \bar{x}) > 0, (y_i - \bar{y}) < 0$

$\Rightarrow (x_i - \bar{x})(y_i - \bar{y}) < 0$

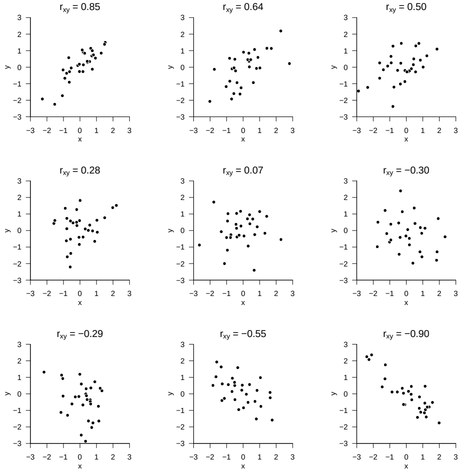
Häufige richtungsgleiche Abweichung der x_i und y_i von ihren Mittelwerten $\Rightarrow$ Positive Korrelation

Häufige richtungsungleiche Abweichung der x_i und y_i von ihren Mittelwerten $\Rightarrow$ Negative Korrelation

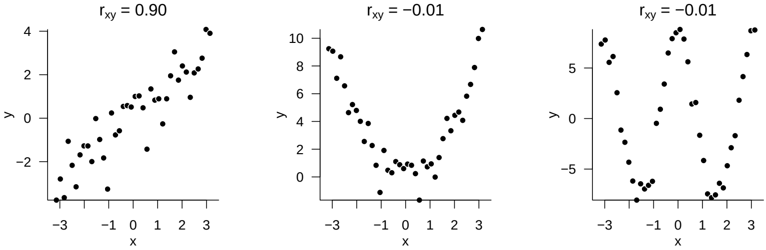
Keine häufigen richtungsgleichen oder -entgegengesetzten Abweichungen $\Rightarrow$ Keine Korrelation

Korrelation

Beispiele



Funktionale Abhängigkeiten und Stichprobenkorrelation



$$y_i = x_i + \varepsilon_i \qquad y_i = x_i^2 + \varepsilon_i \qquad y_i = 8\cos(2x_i) + \varepsilon_i$$

$$\varepsilon_i \sim N(0,1)$$

Korrelation

Theorem (Korrelation und linear-affine Abhängigkeit)

x und y seien zwei Zufallsvariablen mit positiver Varianz. Dann besteht genau dann eine lineare-affine Abhängigkeit der Form

$$y = \beta_0 + \beta_1 x \text{ mit } \beta_0, \beta_1 \in \mathbb{R} \tag{6}$$

zwischen x und y, wenn

$$\rho(x,y) = 1 \text{ oder } \rho(x,y) = -1. \tag{7}$$

Bemerkungen

- Für einen Beweis und eine vertiefte Diskussion verweisen wir auf die BSc Lehreinheiten.
- Die linear-affine Abhängigkeit $y = \beta_0 + \beta_1 x$ impliziert eine linear-affine Abhängigkeit $x = \tilde{\beta}_0 + \tilde{\beta}_1 y$, denn

$$y = \beta_0 + \beta_1 x \Leftrightarrow -\beta_0 + y = \beta_1 x \Leftrightarrow x = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} y \Leftrightarrow x = \tilde{\beta}_0 + \tilde{\beta}_1 y \tag{8}$$

mit

$$\tilde{\beta}_0 = -\frac{\beta_0}{\beta_1} \text{ und } \tilde{\beta}_1 = \frac{1}{\beta_1}. \tag{9}$$

Theorem (Kovarianz und Korrelation bei linear-affinen Transformationen)

ξ und υ seien Zufallsvariablen und es seien $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. Dann gelten

$$\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta) = \alpha\beta\mathbb{C}(\xi, \upsilon) \tag{10}$$

und

$$\rho(\alpha\xi + \beta, \gamma\upsilon + \delta) = \rho(\xi, \upsilon). \tag{11}$$

Bemerkungen

- Wir benötigen diese Aussage im Kontext der Kanonischen Korrelationsanalyse.
- Die Kovarianz zweier Zufallsvariablen ändert sich bei linear-affiner Transformation der Zufallsvariablen.
- Die Korrelation zweier Zufallsvariablen ändert sich bei linear-affiner Transformation der Zufallsvariablen nicht.

Korrelation

Beweis

Es gilt zunächst

$$
\begin{aligned}
\mathbb{C}(\alpha\xi + \beta, c\upsilon + d) &= \mathbb{E}((\alpha\xi + \beta - \mathbb{E}(\alpha\xi + \beta))(\gamma\upsilon + \delta - \mathbb{E}(\gamma\upsilon + \delta))) \\
&= \mathbb{E}((\alpha\xi + \beta - \alpha\mathbb{E}(\xi) - \beta)(\gamma\upsilon + \delta - \gamma\mathbb{E}(\gamma\upsilon) - \delta)) \\
&= \mathbb{E}(\alpha(\xi - \mathbb{E}(\xi))(\gamma(\upsilon - \gamma\mathbb{E}(\upsilon))) \\
&= \mathbb{E}(\alpha\gamma((\xi - \mathbb{E}(\xi))(\upsilon - \gamma\mathbb{E}(\upsilon)))) \\
&= \alpha\gamma\mathbb{C}(\xi, \upsilon)
\end{aligned}
\tag{12}
$$

Also folgt

$$
\begin{aligned}
\rho(\alpha\xi + \beta, \gamma\upsilon + \delta) &= \frac{\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta)}{\sqrt{\mathbb{V}(\alpha\xi + \beta)}\sqrt{\mathbb{V}(\gamma\upsilon + \delta)}} \\
&= \frac{\alpha\gamma\mathbb{C}(\xi, \upsilon)}{\sqrt{\alpha^2\mathbb{V}(\xi)}\sqrt{\gamma^2\mathbb{V}(\upsilon)}} \\
&= \frac{\alpha\gamma\mathbb{C}(\xi, \upsilon)}{\alpha\mathbb{S}(\xi)c\mathbb{S}(\upsilon)} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\mathbb{S}(\xi)\mathbb{S}(\upsilon)} \\
&= \rho(\xi, \upsilon).
\end{aligned}
\tag{13}
$$

Anwendungsszenario Kanonische Korrelation

Therapiegüte als Therapieerfolgsfaktor?

Unabhängige Variablen

- Anzahl Therapiestunden
- Erfahrung Therapeut:in

$\Rightarrow$ Maß für Therapiegüte

Abhängige Variablen

- BDI Score Reduktion
- Glucocorticoid Reduktion

$\Rightarrow$ Maß für Therapieerfolg

Beispieldatensatz Kanonische Korrelation

$i = 1, ..., n$ Patient:innen

y_{1i} BDI Score Reduktion, y_{2i} Glucocorticoid Reduktion, x_{1i} Therapiedauer, x_{2i} Erfahrung Psychotherapeut:in,

x_1i	x_2i	y_1i	y_2i
27.9	7.774	35.5	6.106
15.3	9.347	25.0	3.961
17.4	2.121	19.7	1.716
21.5	6.517	28.8	2.617
28.2	1.256	29.4	1.901
14.0	2.672	17.2	0.872
28.0	3.861	32.9	2.005
28.9	0.134	28.3	4.073
23.2	3.824	25.8	3.918
22.6	8.697	31.3	3.770
11.2	3.403	14.4	2.070
14.1	4.821	18.4	1.999
13.5	5.996	19.1	4.994
23.7	4.935	28.0	2.566
17.7	1.862	20.3	2.086
25.4	8.274	34.8	4.445
20.0	6.685	27.6	3.951
24.4	7.942	31.9	3.851
29.8	1.079	32.2	0.976
17.6	7.237	24.6	1.944

Grundzüge der Kanonischen Korrelationsanalyse

Die Datenvektoren $x_{1i}, \ldots, x_{m_x i}, i = 1, \ldots, n$ werden als u.i.v. Realisierungen eines Zufallsvektors x interpretiert. Die Datenvektoren $y_{1i}, \ldots, y_{m_y i}, i = 1, \ldots, n$ werden als u.i.v. Realisierungen eines Zufallsvektors y interpretiert.

Die "erste kanonische Korrelation" ist die maximale Korrelation von Linearkombinationen von x und y; wir bezeichnen die Linearkombinationen von x und y mit Vektoren $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ mit

$$\xi = a^T x = a_1 x_1 + \cdots + a_{m_x} x_{m_x} \text{ und } \upsilon = b^T y = b_1 y_1 + \cdots + b_{m_y} y_{m_y} \tag{14}$$

ξ und υ sind dann als Linearkombinationen von Zufallsvariablen selbst Zufallsvariablen; Die Korrelation von ξ und υ bezeichnen wir mit $\rho(\xi, \upsilon)$

Wenn die Zufallsvektoren x als unabhängige Variable und y als abhängige Variable interpretiert werden, dann kann $\xi = a^T x$ als "bester Prädiktor" und $\upsilon = b^T y$ als "am besten prädizierbares Kriterium" interpretiert werden. Kanonische Korrelationsanalyse fragt damit nach Parametern $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ für die $\rho(\xi, \upsilon)$ maximal ist.

Für Skalare $\alpha, \beta \in \mathbb{R}$ sind die Korrelationen $\rho(a^T x, b^T y)$ und $\rho((\alpha a^T)x, (\beta b^T)y)$ allerdings identisch (siehe unten). Man sucht deshalb Parameter $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ für die $\rho(\xi, \upsilon)$ maximal ist und für die $a^T x$ und $b^T y$ jeweils eine Varianz von 1 haben, also $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ gilt.

Anders ausgedrückt: Die Varianzen von $a^T x$ und $b^T y$ und die Varianzen von Linearkombinationen von x und y mit beliebigen skalaren Vielfachen von a und b sind im Sinne der ersten Aussage des Theorems zur Kovarianz und Korrelation bei linear-affinen Transformationen verschieden.

Grundzüge der Kanonischen Korrelationsanalyse

Insbesondere gilt einen m_x-dimensionalen Zufallsvektoren x und einen m_y-dimensionalen Zufallsvektor y, $a \in \mathbb{R}^{m_x}$, $b \in \mathbb{R}^{m_y}$, $\alpha, \beta \in \mathbb{R}$ und die Linearkombinationen $\xi := a^T x$ und $\upsilon := b^T y$ basierend auf dem Theorems zur Kovarianz und Korrelation bei linear-affinen allerdings auch, dass

$$\rho(\xi, \upsilon) = \rho(\alpha\xi, \beta\upsilon) \Leftrightarrow \rho(a^T x, b^T y) = \rho(\alpha(a^T x), \beta(b^T y)) \Leftrightarrow \rho(a^T x, b^T y) = \rho((\alpha a^T)x), (\beta b^T)y).$$

Die Korrelation von $a^T x$ und $b^T y$ und die Korrelationen von Linearkombinationen von x und y mit beliebigen skalaren Vielfachen von a und b sind also gleich.

Zur Entwicklung der Kanonischen Korrelationsanalyse folgen wir Mardia, Kent, and Bibby (1979), Kapitel 10. Dabei werden die Zufallsvektoren x und y in einem Zufallsvektor

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \tag{15}$$

zusammengefasst, für den wir durchgängig annehmen, dass $\mathbb{E}(z) = 0_m$ mit $m = m_x + m_y$. Dies entspricht auf der Anwendungsebene der Subtraktion des Stichprobenmittels von den beobachteten Daten vor Durchführung der Kanonischen Korrelationsanalyse

Der mathematische Fokus der Entwicklung nach Mardia, Kent, and Bibby (1979), Kapitel 10 ist auf der Kovarianzmatrix $\mathbb{C}(z)$. Speziell ergeben sich die Kovarianzen von Linearkombinationen von x und y aus Matrixprodukten von $\mathbb{C}(z)$ und es können einige Matrixtheoreme, dieim Folgenden diskutiert werden, auf diese Matrixprodukte angewendet werden. Generell wird in der Entwicklung nach Mardia, Kent, and Bibby (1979), Kapitel 10 ein restringierter Optimierungsansatz mithilfe der Lagrangefunktion zugunsten der Eigenanalyse von Matrixprodukten supprimiert. Für die Entwicklung mit einem Lagrangeansatz, siehe zum Beispiel Anderson (2003), Kapitel 12.

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Definition (Symmetrische Quadratwurzel einer Matrix)

$A \in \mathbb{R}^{m \times m}$ sei eine invertierbare symmetrische Matrix mit positiven Eigenwerten. Dann sind für $r \in \mathbb{N}^0$ und $s \in \mathbb{N}$ die rationalen Potenzen von A einer orthonormalen Matrix $Q \in \mathbb{R}^{m \times m}$ der Eigenvektoren von A und einer Diagonalmatrix $\Lambda = \text{diag}(\lambda_i) \in \mathbb{R}^{m \times m}$ der zugehörigen Eigenwerte $\lambda_1, ..., \lambda_m$ von A definiert als

$$A^{r/s} = Q\Lambda^{r/s}Q^T \text{ mit } \Lambda^{r/s} = \text{diag}\left(\lambda_i^{r/s}\right). \tag{16}$$

Der Spezialfall $r := 1, s := 2$ wird als symmetrische Quadratwurzel von A bezeichnet und hat die Form

$$A^{1/2} = Q\Lambda^{1/2}Q^T \text{ mit } \Lambda^{1/2} = \text{diag}\left(\lambda_i^{1/2}\right). \tag{17}$$

Bemerkungen

- Offenbar gilt

$$\left(A^{1/2}\right)^2 = Q\Lambda^{1/2}Q^TQ\Lambda^{1/2}Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = Q\Lambda Q^T = A. \tag{18}$$

- Weiterhin gilt

$$\left(A^{-1/2}\right)^2 = Q\Lambda^{-1/2}Q^TQ\Lambda^{-1/2}Q^T = Q\Lambda^{-1}Q^T = A^{-1}. \tag{19}$$

  Die vorletzte Gleichung mag überraschen, aber es gilt ja zum Beispiel

$$4^{-1/2} \cdot 4^{-1/2} = \frac{1}{\sqrt{4}} \cdot \frac{1}{\sqrt{4}} = \frac{1}{4} = 4^{-1}. \tag{20}$$

Theorem (Eigenwerte und Eigenvektoren von Matrixprodukten)

Für $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times n}$ sind die Eigenwerte von $AB \in \mathbb{R}^{n \times n}$ und $BA \in \mathbb{R}^{m \times m}$ gleich. Weiterhin gilt, dass für einen Eigenvektor v zu einem von Null verschiedenen Eigenwert λ von AB $w := Bv$ ein Eigenvektor von BA ist.

Bemerkungen

- Für einen Beweis siehe Mardia, Kent, and Bibby (1979), S. 468.

```
A   = matrix(1:6, nrow = 2,  byrow = T)          # Matrix A \in \mathbb{R}^{2 x 3}
B   = matrix(1:6, ncol = 2,  byrow = T)          # Matrix B \in \mathbb{R}^{3 x 2}
EAB = eigen(A %*% B)                             # Eigenanalyse von AB \in \mathbb{R}^{2 \times 2}
EBA = eigen(B %*% A)                             # Eigenanalyse von BA \in \mathbb{R}^{3 \times 3}
w   = B %*% EAB$vectors[,1]                      # Eigenvektor von BA
cat("Eigenwerte von AB :"  , EAB$values[1:2],
    "\nEigenwerte von BA :", EBA$values[1:2],
    "\nBAw mit w = Bv    :", B %*% A %*% w,
    "\nlw mit w = Bv     :", EBA$values[1] * w)
```

```
> Eigenwerte von AB : 85.6 0.421
> Eigenwerte von BA : 85.6 0.421
> BAw mit w = Bv    : -191 -417 -642
> lw mit w = Bv     : -191 -417 -642
```

Theorem (Eigenwert und Eigenvektor eines Matrixvektorprodukts)

Für $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{p \times n}$, $a \in \mathbb{R}^m$ und $b \in \mathbb{R}^p$ gilt, dass der einzige von Null verschiedene Eigenwert von $Aab^TB \in \mathbb{R}^{n \times n}$ gleich b^TBAa mit zugehörigem Eigenvektor Aa ist.

Bemerkungen

- Für einen Beweis siehe Mardia, Kent, and Bibby (1979), S. 468.

```
A      = matrix(1:6, nrow = 2,  byrow = T)      # Matrix A \in \mathbb{R}^{2 x 3}
B      = matrix(1:8, ncol = 2,  byrow = T)      # Matrix B \in \mathbb{R}^{4 x 2}
a      = matrix(1:3, nrow = 3,  byrow = T)      # Vektor a \in \mathbb{R}^{3 x 1}
b      = matrix(1:4, nrow = 4,  byrow = T)      # Vektor b \in \mathbb{R}^{4 x 1}
EAabTB = eigen(A %*% a %*% t(b) %*% B)          # Eigenanalyse von Aab^TB \in \mathbb{R}^{4 x 4}
cat("Eigenwerte von AabTB :", EAabTB$values,
    "\nbTBAa                :", t(b) %*% B %*% A %*% a,
    "\nAa                   :", A %*% a,
    "\n(AabTB)Aa            :",(A %*% a %*% t(b) %*% B) %*% A %*% a,            # Mv
    "\n(bTBAa)Aa            :",as.vector((t(b) %*% B %*% A %*% a)) * (A %*% a)) # = \lambda v
```

```
> Eigenwerte von AabTB : 2620 0
> bTBAa                : 2620
> Aa                   : 14 32
> (AabTB)Aa            : 36680 83840
> (bTBAa)Aa            : 36680 83840
```

Theorem (Maximierung quadratischer Formen mit Nebenbedingungen)

$A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{m \times m}$ p.d. seien symmetrische Matrizen und λ_1 sei der größte Eigenwert von $B^{-1}A$ mit assoziiertem Eigenvektor $v_1 \in \mathbb{R}^m$. Dann ist λ_1 eine Lösung des Optimierungsproblems

$$\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{21}$$

Bemerkungen

- Das Theorem ist direkt durch die kanonische Korrelationsanalyse motiviert.
- $\max_x f(x)$ ist das Maximum einer Funktion f, also der Wert der Funktion an der Maximumstelle x
- $\arg\max_x f(x)$ ist die Maximumstelle einer Funktion, also ein Wert in der Definitionsmenge von f.
- Nach Wortlaut des Theorems gilt also für die Funktion

$$f : \mathbb{R}^m \to \mathbb{R}, x \mapsto f(x) := x^T Ax, \tag{22}$$

dass

$$v_1 = \arg\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1 \tag{23}$$

und dass

$$\lambda_1 = \max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{24}$$

Beweis

$B^{1/2}$ sei die symmetrische Quadratwurzel von B und es sei

$$y := B^{1/2}x \Leftrightarrow x = B^{-1/2}y \tag{25}$$

Dann kann mit der symmetrischen Matrix

$$K := B^{-1/2}AB^{-1/2} \in \mathbb{R}^{m\times m} \tag{26}$$

das Optimierungsproblem (21) geschrieben werden als

$$\max_y y^T Ky \text{ unter der Nebenbedingung } y^T y = 1. \tag{27}$$

Dies gilt, weil

$$\max_x x^T Ax \Leftrightarrow \max_y \left(B^{-1/2}y\right)^T A \left(B^{-1/2}y\right) \Leftrightarrow \max_y y^T B^{-1/2}AB^{-1/2}y \Leftrightarrow \max_y y^T Ky \tag{28}$$

und

$$x^T Bx = 1 \Leftrightarrow y^T B^{-1/2}BB^{-1/2}y = 1 \Leftrightarrow y^T y = 1. \tag{29}$$

Weil K eine symmetrische Matrix ist, existiert die Orthonormalzerlegung (vgl. (2) Matrizen)

$$K = Q\Lambda Q^T, \tag{30}$$

wobei die Spalten der orthogonalen Matrix Q die Eigenvektoren von K und die Diagonalemente von Λ die zugehörigen Eigenwerte von K sind.

Beweis (fortgeführt)

Mit der orthogonalen Matrix Q aus obiger Orthornomalzerlegung sei nun

$$z := Q^T y \Leftrightarrow y := Qz. \tag{31}$$

Dann kann das Optimierungsproblem (27) geschrieben werden als

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \text{ unter der Nebenbedingung } z^T z = 1, \tag{32}$$

weil

$$\max_y y^T K y \Leftrightarrow \max_z (Qz)^T K(Qz) \Leftrightarrow \max_z z^T Q^T Q \Lambda Q^T Q z \Leftrightarrow \max_z z^T \Lambda z \Leftrightarrow \max_z \sum_{i=1}^m \lambda_i z_i^2 \tag{33}$$

und

$$y^T y = 1 \Leftrightarrow (Qz)^T Qz = 1 \Leftrightarrow z^T Q^T Q z = 1 \Leftrightarrow z^T z = 1. \tag{34}$$

Beweis (fortgeführt)

Die Eigenwerte von K seien nun absteigend sortiert, also $\lambda_1 \geq \cdots \geq \lambda_m$. Dann gilt für das Optimierungsproblem (32), dass

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_1, \tag{35}$$

weil

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \max_z \sum_{i=1}^m \lambda_1 z_i^2 = \lambda_1 \max_z \sum_{i=1}^m z_i^2 = \lambda_1 \tag{36}$$

wobei sich die letzte Gleichung aus der Nebenbedingung $z^T z = 1$ ergibt. Schließlich gilt

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 = \lambda_1, \tag{37}$$

für $z := e_1 = (1, 0, ..., 0)^T$. Zusammenfassend heißt das, dass $z = e_1$ eine Lösung des Optimierungsproblem (32) ist und das λ_1 das entsprechende Maximum ist.

Beweis (fortgeführt)

Damit ergibt sich aber sofort, dass dann

$$y = Qz = Qe_1 = q_1 \text{ und } x = B^{-1/2} q_1 \tag{38}$$

Lösungen der äquivalenten Optimierungsprobleme (27) und (21), respektive, sind. Nach Konstruktion ist q_1 ein Eigenvektor von $B^{-1/2}AB^{-1/2}$ und nach obigem Theorem zu Eigenwerten und Eigenvektoren von Matrixprodukten damit auch ein Eigenvektor von

$$B^{-1/2}B^{-1/2}A = B^{-1}A \tag{39}$$

und die zugehörigen Eigenwerte sind gleich. Damit aber folgt, dass der größte Eigenwert von $B^{-1}A$ und sein assoziierter Eigenvektor eine Lösung von

$$\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{40}$$

ist. $\square$

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Theorem (Kovarianzmatrizen von Zufallsvektoren)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \tag{41}$$

ein $m_x + m_y$-dimensionaler Zufallsvektor und sein Erwartungswertvektor, respektive. Dann kann die $m \times m$ Kovarianzmatrix z geschrieben werden als

$$\mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{42}$$

wobei

$$\begin{aligned} \Sigma_{xx} &:= \mathbb{E}\left(xx^T\right) \in \mathbb{R}^{m_x \times m_x} \\ \Sigma_{xy} &:= \mathbb{E}\left(xy^T\right) \in \mathbb{R}^{m_x \times m_y} \\ \Sigma_{yx} &:= \mathbb{E}\left(yx^T\right) \in \mathbb{R}^{m_y \times m_x} \\ \Sigma_{yy} &:= \mathbb{E}\left(yy^T\right) \in \mathbb{R}^{m_x \times m_y} \end{aligned} \tag{43}$$

Beweis

Nach Definition der Kovarianzmatrix eines Zufallsvektors gilt

$$
\begin{aligned}
\mathbb{C}(z) &= \mathbb{E}\left((z - \mathbb{E}(z))(z - \mathbb{E}(z))^T\right) \\
&= \mathbb{E}\left((z - 0_m)(z - 0_m)^T\right) \\
&= \mathbb{E}\left(zz^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} x^T & y^T \end{pmatrix}\right) \\
&= \mathbb{E}\left(\begin{pmatrix} xx^T & xy^T \\ yx^T & yy^T \end{pmatrix}\right) \\
&= \begin{pmatrix} \mathbb{E}\left(xx^T\right) & \mathbb{E}\left(xy^T\right) \\ \mathbb{E}\left(yx^T\right) & \mathbb{E}\left(yy^T\right) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}
\end{aligned}
\tag{44}
$$

□

Theorem (Linearkombinationen von Zufallsvektorpartitionen)

Es sei

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(x) = 0_m \text{ und } \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{45}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswertvektor und seine Kovarianzmatrix, respektive. Weiterhin seien für $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ die Zufallsvariablen

$$\xi := a^T x \text{ und } \upsilon := b^T y \tag{46}$$

als Linearkombinationen der Komponenten von x und y definiert. Dann gelten

(1) $\mathbb{V}(\xi) = a^T \Sigma_{xx} a$

(2) $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b$

(2) $\rho(\xi, \upsilon) = a^T \Sigma_{xy} b$, wenn $\mathbb{V}(\xi) = 1$ und $\mathbb{V}(\upsilon) = 1$.

Bemerkungen

- Die Varianz der Zufallsvariable $a^T x$ ergibt sich als “doppelte Linearkombination” von Σ_{xx}.
- Die Varianz der Zufallsvariable $b^T y$ ergibt sich als “doppelte Linearkombination” von Σ_{yy}.
- Die Korrelation der Zufallsvariablen $a^T x$ und $b^T y$ ergibt sich “doppellte Linearkombination” von Σ_{xy}.

Beweis von (1) und (2)

Wir betrachten zunächst die Varianz von ξ. Mit dem Varianzverschiebungssatz (vgl. Erwartungswert, Varianz, Kovarianz) gilt

$$
\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}(\xi\xi) - \mathbb{E}(\xi)\mathbb{E}(\xi) \\
&= \mathbb{E}\left((a^Tx)(a^Tx)\right) - \mathbb{E}\left(a^Tx\right)\mathbb{E}\left(a^Tx\right) \\
&= \mathbb{E}\left((a^Tx)(a^Tx)^T\right) - \mathbb{E}\left(a^Tx\right)\mathbb{E}\left(a^Tx\right) \\
&= \mathbb{E}\left(a^Txx^Ta\right) - \mathbb{E}\left(a^Tx\right)\mathbb{E}\left(a^Tx\right) \\
&= a^T\mathbb{E}\left(xx^T\right)a - a^T\mathbb{E}(x)a^T\mathbb{E}(x) \\
&= a^T\mathbb{E}\left(xx^T\right)a - a^T0_{m_x}a^T0_{m_x} \\
&= a^T\Sigma_{xx}a.
\end{aligned}
\tag{47}
$$

Der Beweis zur Varianz von υ folgt dann analog.

Modellformulierung

Beweis von (3)

Mit der Definition der Korrelation von Zufallsvariablen und mit $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ und dem Kovarianzverschiebungssatz (vgl. Erwartungswert, Varianz, Kovarianz) gilt

$$
\begin{aligned}
\rho(\xi, \upsilon) &= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(\upsilon)}} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{1}\sqrt{1}} \\
&= \mathbb{C}(\xi, \upsilon) \\
&= \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) \\
&= \mathbb{E}\left((a^Tx)(b^Ty)\right) - \mathbb{E}(a^Tx)\mathbb{E}(b^Ty) \\
&= \mathbb{E}\left((a^Tx)(b^Ty)^T\right) - \mathbb{E}(a^Tx)\mathbb{E}(b^Ty) \\
&= \mathbb{E}\left(a^Txy^Tb\right) - \mathbb{E}(a^Tx)\mathbb{E}(b^Ty) \\
&= a^T\mathbb{E}\left(xy^T\right)b - a^T\mathbb{E}(x)b^T\mathbb{E}(y) \\
&= a^T\mathbb{E}\left(xy^T\right)b - a^T0_{m_x}b^T0_{m_y} \\
&= a^T\Sigma_{xy}b.
\end{aligned} \tag{48}
$$

□

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Definition (Kanonische Koeffizientenvektoren, Variate, Korrelationen)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \text{ und } \mathbb{C}(z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{49}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswert und seine Kovarianzmatrix, respektive. Weiterhin sei

$$K := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{50}$$

mit der Singulärwertzerlegung

$$K = A \Lambda B^T, \tag{51}$$

wobei

$$A := \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \end{pmatrix} \in \mathbb{R}^{m_x \times m_y} \text{ und } B := \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix} \in \mathbb{R}^{m_y \times m_y} \tag{52}$$

die orthogonalen Matrix der Eigenvektoren von KK^T und die orthogonale Matrix der Eigenvektoren von $K^T K$, respektive, bezeichnen und

$$\Lambda := \operatorname{diag}\left(\lambda_1^{1/2}, ..., \lambda_k^{1/2}\right) \in \mathbb{R}^{m_y \times m_y}, \tag{53}$$

die Diagonalmatrix der Quadratwurzeln der zugehörigen Eigenvektoren bezeichnet. Schließlich seien für $i = 1, ..., k$

$$a_i := \Sigma_{xx}^{-1/2} \alpha_i \in \mathbb{R}^{m_x} \text{ und } b_i := \Sigma_{yy}^{-1/2} \beta_i \in \mathbb{R}^{m_y}. \tag{54}$$

Dann heißen für $i = 1, ..., k$

(1) $a_i \in \mathbb{R}^{m_x}$ und $b_i \in \mathbb{R}^{m_y}$ die *$i$ten kanonischen Koeffizientenvektoren*,

(2) die Zufallsvektoren $\xi_i := a_i^T x$ und $\upsilon_i := b_i^T y$ die $i$ten *$i$ten kanonischen Variaten* und

(3) $\rho_i := \lambda_i^{1/2}$ die *$i$te kanonische Korrelation*.

Theorem (Eigenschaften kanonischer Korrelationen und Variaten)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \text{ und } \mathbb{C}(z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{55}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswert und seine Kovarianzmatrix, respektive. Weiterhin seien für $i = 1, ..., k$ die kanonischen Koeffizientenvektoren a_i, b_i, die kanonischen Variaten ξ, υ_i und die kanonischen Korrelationen ρ_i definiert wie oben. Dann gilt, dass für $1 \le r \le k$ das Maximum des $r$ten restringierten Optimierungsproblems

$$\phi_r = \max_{a,b} a^T \Sigma_{xy} b \tag{56}$$

unter den Nebenbedingungen

$$a^T \Sigma_{xx} a = 1, \quad b^T \Sigma_{yy} b = 1, \quad a_i^T \Sigma_{xx} a = 0 \text{ für } i = 1, ..., r-1 \tag{57}$$

(1) den Wert $\phi_r = \rho_r$ hat und (2) bei $a = a_r$ und $b = b_r$ angenommen wird.

Bemerkungen

- ϕ_1 ist die größtmögliche Korrelation von $\xi = a^T x$ und $\upsilon = b^T y$ unter den Nebenbedingungen
  - $\mathbb{V}(\xi) = a^T \Sigma_{xx} a = 1$ und $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b = 1$
- ϕ_r ist die größtmögliche Korrelation von $\xi = a^T x$ und $\upsilon = b^T y$ unter den Nebenbedingungen
  - $\mathbb{V}(\xi) = a^T \Sigma_{xx} a = 1$ und $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b = 1$
  - $\mathbb{C}(\xi_i, \xi) = a_i^T \Sigma_{xx} a = 0$ für die ersten $i = 1, ..., r-1$ kanonischen Variaten ξ_i

Modellformulierung

Beweis

Wir betrachten das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a,b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ b^T \Sigma_{yy} b = 1,\ a_i^T \Sigma_{xx} a = 0, i = 1, \ldots, r-1 \tag{58}$$

Wir folgen Mardia, Kent, and Bibby (1979), S. 284 und gehen schrittweise vor, d.h. wir lösen das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a} \left(\max_{b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N.} b^T \Sigma_{yy} b = 1 \right) \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ a_i^T \Sigma_{xy} a = 0, i = 1, \ldots, r-1 \tag{59}$$

von innen nach außen.

Schritt (1)

Wir wählen wir zunächst ein festes $a \in \mathbb{R}^m$ und betrachten das restringierte Optimierungsproblem

$$\max_{b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \tag{60}$$

Dieses Optimierungsproblem kann geschrieben werden als

$$\max_{b} b^T \Sigma_{yx} a a^T \Sigma_{xy} b \text{ u.d.N. } b^T \Sigma_{yy} b = 1, \tag{61}$$

weil gilt, dass

$$\left(a^T \Sigma_{xy} b\right)^2 = \left(a^T \Sigma_{xy} b\right) \left(a^T \Sigma_{xy} b\right) = \left(a^T \Sigma_{xy} b\right)^T a^T \Sigma_{xy} b = b^T \Sigma_{yx} a a^T \Sigma_{xy} b. \tag{62}$$

Beweis (fortgeführt)

Das Optimierungsproblem (61) kann nun mithilfe des Theorems zur Maximierung quadratischer Formen mit Nebenbedingungen gelöst werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yx} a a^T \Sigma_{xy} \text{ und } B := \Sigma_{yy}. \tag{63}$$

Dann hat (61) die Form

$$\max_b b^T A b \text{ unter der Nebenbedingung } b^T B b = 1, \tag{64}$$

Das Maximum von (64) entspricht nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen dem größten Eigenwert von

$$B^{-1} A = \Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy} \tag{65}$$

Der größte Eigenwert von $\Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy}$ wiederum kann mithilfe des Theorems zum Eigenwert und Eigenvektor eines Matrixvektorprodukts bestimmt werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yy}^{-1} \Sigma_{yx}, \quad b := a, \quad B := \Sigma_{xy} \tag{66}$$

und erhalten für den betreffenden Eigenwert

$$\lambda_a = b^T B A a = a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a. \tag{67}$$

als Lösung (Maximum) des restringierten Optimierungsproblems

$$\max_b \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \tag{68}$$

Modellformulierung

Beweis (fortgeführt)

Schritt (2)

Basierend auf Schritt (1) verbleibt die Lösung des restringierten Optimierungsproblem

$$\phi_r^2 = \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ a_i^T \Sigma_{xx} a = 0, i = 1, ..., r-1 \tag{69}$$

Dazu halten wir zunächst fest, dass (69) mit den Definitionen von α_i und K in der Definition der Kanonischen Koeffizientenvektoren, Variaten, und Korrelationen geschrieben werden kann als

$$\phi_r^2 = \max_\alpha \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1,\ \alpha_i^T \alpha = 0, i = 1, ..., r-1, \tag{70}$$

denn

$$\begin{aligned}
\phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0 \Leftrightarrow \\
\phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 1, \alpha_i^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 0 \\
\phi_r^2 &= \max_a \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\
\phi_r^2 &= \max_a \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\
\phi_r^2 &= \max_a \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0
\end{aligned} \tag{71}$$

Modellformulierung

Beweis (fortgeführt)

Dabei sind nach der betreffenden Definition die α_i die Eigenvektoren von KK^T mit den $i = 1, ..., r-1$ größten Eigenwerten. Nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen ist die Lösung von (70) der größte Eigenwert von KK^T mit seinem assoziierten Eigenvektor. Die Nebenbedingung $\alpha_i^T \alpha = 0$ schränkt diese Wahl auf den rt-größten Eigenwert und seinen assoziierten Eigenvektor α_r ein. Mit der Definition von Eigenwerten und Eigenvektoren gilt also

$$\phi_r^2 = \alpha_r^T KK^T \alpha_r = \alpha_r^T \lambda_r \alpha_r = \lambda_r \alpha_r^T \alpha_r = \lambda_r. \tag{72}$$

Wir haben also gezeigt, dass das restringierte Optimierungsproblem des Theorems den Maximumwert $\phi_r = \lambda_r^{1/2}$ hat. Es bleibt zu zeigen, dass dieser Maximumwert für a_r und b_r angenommen wird.

Schritt (3)

Einsetzen von a_r und b_r in $a^T \Sigma_{xy} b$ ergibt mit

$$K = A\Lambda B^T \Leftrightarrow KB = A\Lambda B^T B \Leftrightarrow KB = A\Lambda \Leftrightarrow K\beta_r = \alpha_r \lambda_r^{1/2} \tag{73}$$

dass

$$a_r^T \Sigma_{xy} b_r = \alpha_r^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \beta_r = \alpha_r^T K \beta_r = \alpha_r^T \alpha_r \lambda_r^{1/2} = \rho_r \tag{74}$$

Also nimmt $a^T \Sigma_{xy} b$ bei a_r und b_r seinen restringierten Maximalwert λ_r an.

$\square$

Simulationsbeispiel

Wir betrachten das Beispiel (vgl. Uurtio et al. (2018))

$$p(x) = N(x; 0_4, I_4) \text{ und } p(y|x) = N(y; LX, G) \tag{75}$$

mit

$$L := \begin{pmatrix} 0.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 \end{pmatrix} \text{ und } G := \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \tag{76}$$

Hier gilt offenbar $m_x = 4, m_y = 3, m = 7$ und

$$\begin{aligned} y_1 &= x_3 + \varepsilon_1 \\ y_2 &= x_1 + \varepsilon_2 \\ y_3 &= -x_4 + \varepsilon_3 \end{aligned} \tag{77}$$

mit

$$x_1 \sim N(0,1), x_3 \sim N(0,1), x_4 \sim N(0,1) \tag{78}$$

und

$$\varepsilon_1 \sim N(0, 0.2), \varepsilon_2 \sim N(0, 0.4), \varepsilon_3 \sim N(0, 0.3) \tag{79}$$

Simulationsbeispiel

Mit dem Theorem zu gemeinsamen Normalverteilungen (vgl. Einheit (3) Matrizen) ergibt sich, dass

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N(0_7, \Sigma) \tag{80}$$

mit

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{81}$$

wobei

$$\Sigma_{xx} = I_4, \quad \Sigma_{xy} = L^T, \quad \Sigma_{yx} = L \text{ und } \Sigma_{yy} = G + LL^T. \tag{82}$$

Explizit ergibt sich also

$$\Sigma = \begin{pmatrix} I_4 & L^T \\ L & G + LL^T \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.2 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & 1.3 \end{pmatrix} \tag{83}$$

Modellformulierung

Simulationsbeispiel

```
# R Pakete für Matrizenrechnung
library(matlib)
library(expm)

# Modellparameter
L = matrix(c(0,0,1, 0,
             1,0,0, 0,
             0,0,0,-1),
           nrow  = 3,
           byrow = T)
G = diag(c(0.2,0.4,0.3))

# Kovarianzmatrixpartition
Sigma_xx = diag(4)
Sigma_xy = t(L)
Sigma_yx = L
Sigma_yy = G + L %*% t(L)
Sigma    = rbind(cbind(Sigma_xx, Sigma_xy), cbind(Sigma_yx, Sigma_yy))
print(Sigma)
```

```
>      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
> [1,]    1    0    0    0  0.0  1.0  0.0
> [2,]    0    1    0    0  0.0  0.0  0.0
> [3,]    0    0    1    0  1.0  0.0  0.0
> [4,]    0    0    0    1  0.0  0.0 -1.0
> [5,]    0    0    1    0  1.2  0.0  0.0
> [6,]    1    0    0    0  0.0  1.4  0.0
> [7,]    0    0    0   -1  0.0  0.0  1.3
```

Modellformulierung

Simulationsbeispiel

```
# Evaluation der iten kanonischen Koeffizientenvektoren und Korrelationen
K      = sqrtm(inv(Sigma_xx)) %*% Sigma_xy %*% sqrtm(inv(Sigma_yy)) # K
ALB    = svd(K)                                                     # K = A\LambdaV
A      = ALB$u                                                      # A
Lambda = ALB$d                                                      # Lambda
B      = ALB$v                                                      # B
rho    = Lambda                                                     # \rho_i = \lambda_i^{1/2}
a      = sqrtm(inv(Sigma_xx)) %*% A                                 # a_i = \Sigma_{xx}^{-1/2}\alpha_i
b      = sqrtm(inv(Sigma_yy)) %*% B                                 # b_i = \Sigma_{yy}^{-1/2}\beta_i
```

Die kanonische Korrelationen und kanonischen Koeffizientenvektoren ergeben sich zu

```
> rho_1 =  0.913 , a_1^T = ( 0 0 -1 0 ), b_1^T = ( -0.913 0 0 )
> rho_2 =  0.877 , a_2^T = ( 0 0 0 1 ) , b_2^T = ( 0 0 -0.877 )
> rho_3 =  0.845 , a_3^T = ( -1 0 0 0 ), b_3^T = ( 0 -0.845 0 )
```

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Definition (Schätzer kanonischer Korrelationen und Koeffizientenvektoren)

Für $i = 1, ..., n$ seien

$$z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \text{ mit } \mathbb{E}(z_i) := 0_m \text{ und } \mathbb{C}(z_i) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{84}$$

unabhängig und identisch verteilte m-dimensionale partitionierte Zufallsvektoren sowie ihr Erwartungswert und ihre Kovarianzmatrix, respektive, und

$$C := \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{85}$$

sei ihre Stichprobenkovarianzmatrix. Dann sind für $i = 1, ..., k := \min\{m_x, m_y\}$

$$\hat{a}_i := C_{xx}^{-1/2} \hat{\alpha}_i \in \mathbb{R}^{m_x}, \quad \hat{b}_i := C_{yy}^{-1/2} \hat{\beta}_i \in \mathbb{R}^{m_y} \text{ und } \hat{\rho}_i := \hat{\lambda}_i^{1/2} \tag{86}$$

Schätzer der $i$ten kanonischen Koeffizientenvektoren und kanonischen Korrelationen, respektive. Dabei sind mit

$$\hat{K} := C_{xx}^{-1/2} C_{xy} C_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{87}$$

$\hat{\alpha}_i$ und $\hat{\lambda}_i$ der $i$te Eigenvektor und sein zugehöriger Eigenwert von $\hat{K}\hat{K}^T$ und $\hat{\beta}_i$ der entsprechende Eigenkvektor von $\hat{K}^T\hat{K}$.

Bemerkungen

- Zur Modellschätzung wird $\mathbb{C}(z)$ also durch C ersetzt.

Simulationsbeispiel

```r
# R Pakete
library(MASS)
library(matlib)
library(expm)

# Modellparameter
m_x      = 4
m_y      = 3
k        = min(m_x,m_y)
L        = matrix(c(0,0,1,0,1,0,0,0,0,0,0,-1), nrow = 3,byrow = 3)
G        = diag(c(0.2,0.4,0.3))
Sigma_xx = diag(4)
Sigma_xy = t(L)
Sigma_yx = L
Sigma_yy = G + L %*% t(L)
Sigma    = rbind(cbind(Sigma_xx, Sigma_xy), cbind(Sigma_yx, Sigma_yy))
K        = sqrtm(inv(Sigma_xx)) %*% Sigma_xy %*% sqrtm(inv(Sigma_yy))
ALB      = svd(K)
A        = ALB$u
Lambda   = ALB$d
B        = ALB$v
rho      = Lambda
a        = sqrtm(inv(Sigma_xx)) %*% A
b        = sqrtm(inv(Sigma_yy)) %*% B
```

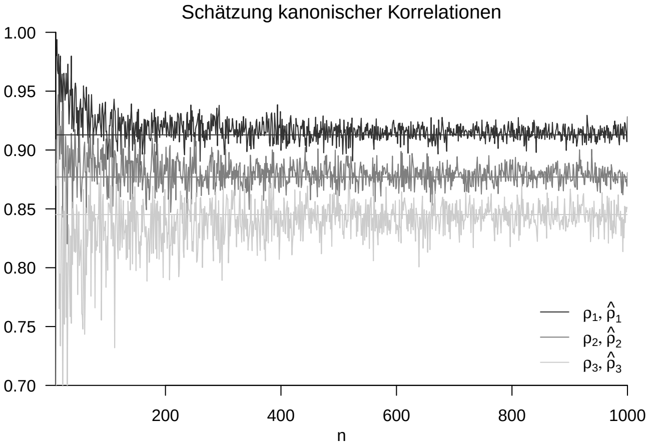
Modellschätzung

Simulationsbeispiel

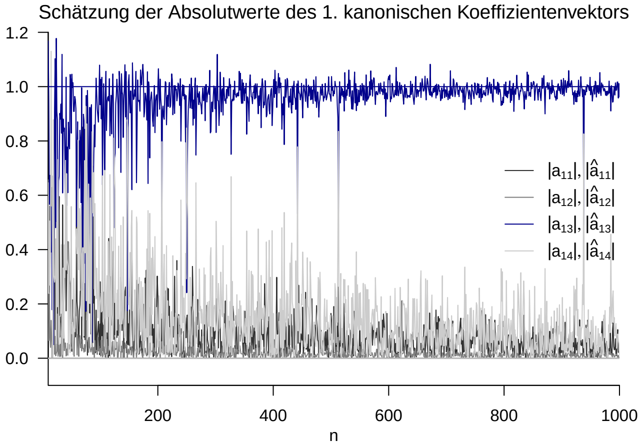
```
# Simulationen
n       = 1e1:1e3
rho_hat = matrix(rep(NaN, length(n)*k)   , nrow = k)
a_1_hat = matrix(rep(NaN, length(n)*m_x), nrow = m_x)
for(i in 1:length(n)){

    # Datengeneration
    Y           = t(mvrnorm(n[i],rep(0, m_x+m_y),Sigma))
    I_n         = diag(n[i])
    J_n         = matrix(rep(1,n[i]^2), nrow = n[i])

    # Stichprobenkovarianzmatrixpartition
    C           = (1/(n[i]-1))*(Y %*% (I_n-(1/n[i])*J_n) %*% t(Y))
    C_xx        = C[1:m_x,1:m_x]
    C_xy        = C[1:m_x,(m_x+1):(m_x+m_y)]
    C_yx        = C[(m_x+1):(m_x+m_y),1:m_x]
    C_yy        = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]

    # Kanonische Korrelationsanalyse
    K_hat         = sqrtm(inv(C_xx)) %*% C_xy %*% sqrtm(inv(C_yy))
    ALB_hat       = svd(K_hat)
    A_hat         = ALB_hat$u
    Lambda_hat    = ALB_hat$d
    B_hat         = ALB_hat$v
    a_hat         = sqrtm(inv(C_xx)) %*% A_hat
    b_hat         = sqrtm(inv(C_yy)) %*% B_hat
    rho_hat[,i]   = as.matrix(Lambda_hat)
    a_1_hat[,i]   = a_hat[,1]
}
```

Modellschätzung

Simulationsbeispiel



Simulationsbeispiel



Anwendungsbeispiel

Therapiegüte als Therapieerfolgsfaktor?

Unabhängige Variablen

- Anzahl Therapiestunden
- Erfahrung Therapeut:in

$\Rightarrow$ Maß für Therapiegüte

Abhängige Variablen

- BDI Score Reduktion
- Glucocorticoid Reduktion

$\Rightarrow$ Maß für Therapieerfolg

Anwendungsbeispiel

$i = 1, ..., n$ Patient:innen

y_{1i} BDI Score Reduktion, y_{2i} Glucocorticoid Reduktion, x_{1i} Therapiedauer, x_{2i} Erfahrung Psychotherapeut:in,

x_1i	x_2i	y_1i	y_2i
27.9	7.774	35.5	6.106
15.3	9.347	25.0	3.961
17.4	2.121	19.7	1.716
21.5	6.517	28.8	2.617
28.2	1.256	29.4	1.901
14.0	2.672	17.2	0.872
28.0	3.861	32.9	2.005
28.9	0.134	28.3	4.073
23.2	3.824	25.8	3.918
22.6	8.697	31.3	3.770
11.2	3.403	14.4	2.070
14.1	4.821	18.4	1.999
13.5	5.996	19.1	4.994
23.7	4.935	28.0	2.566
17.7	1.862	20.3	2.086
25.4	8.274	34.8	4.445
20.0	6.685	27.6	3.951
24.4	7.942	31.9	3.851
29.8	1.079	32.2	0.976
17.6	7.237	24.6	1.944

Modellschätzung

Anwendungsbeispiel

```
# libraries
library(expm)
library(matlib)

# Datenpräprozessierung
fname       = file.path(getwd(), "7_Kanonische_Korrelationsanalyse.csv")
D           = read.table(fname, sep = ",", header = TRUE)
x           = as.matrix(cbind(D$x_1i, D$x_2i))
y           = as.matrix(cbind(D$y_1i, D$y_2i))
n           = nrow(x)
m_x         = ncol(x)
m_y         = ncol(y)
Y           = t(cbind(x,y))

# Stichprobenkovarianzmatrixpartition
I_n         = diag(n)
J_n         = matrix(rep(1,n^2), nrow = n)
C           = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))
C_xx        = C[1:m_x,1:m_x]
C_xy        = C[1:m_x,(m_x+1):(m_x+m_y)]
C_yx        = C[(m_x+1):(m_x+m_y),1:m_x]
C_yy        = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]

# Kanonische Korrelationsanalyse
K_hat       = sqrtm(inv(C_xx)) %*% C_xy %*% sqrtm(inv(C_yy))
ALB_hat     = svd(K_hat)
A_hat       = ALB_hat$u
Lambda_hat  = ALB_hat$d
B_hat       = ALB_hat$v
a_hat       = sqrtm(inv(C_xx)) %*% A_hat
b_hat       = sqrtm(inv(C_yy)) %*% B_hat
rho_hat     = as.matrix(Lambda_hat)

print(ALB_hat)
```

```
> $d
> [1] 0.995 0.501
>
> $u
>        [,1]   [,2]
> [1,] -0.897 -0.443
> [2,] -0.443  0.897
```

Anwendungsbeispiel

Kanonische Korrelationsanalyse mir R's `cancor()` Funktion

```
# Datenpräprozessierung
fname      = file.path(getwd(), "7_Kanonische_Korrelationsanalyse.csv")
D          = read.table(fname, sep = ",", header = TRUE)
x          = as.matrix(cbind(D$x_1i, D$x_2i))
y          = as.matrix(cbind(D$y_1i, D$y_2i))
cca        = cancor(x,y)
```

```
> rho_hat_1 :  0.995
> rho_hat_2 :  0.501
```

Modellschätzung

Anwendungsbeispiel

Die geschätzte maximale Korrelation von Linearkombinationen von (x_1, x_2) und (y_1, y_2) ist 0.99.

- (x_1, x_2) und (y_1, y_2) sind multivariat also "hochgradig" korreliert.

Basierend auf der simulationsvalidierten Schätzung ergibt sich

- $\xi = 0.16x_1 + 0.17x_2$ als "bester Prädiktor"
- $\upsilon = 0.15y_1 + 0.05y_2$ als "am besten prädizierbares Kriterium"

"Therapiedauer" und "Therapeut:innenerfahrung" scheinen zur bestmöglichen Prädiktion der Therapiegüte also in etwa gleichbedeutend, bei dem bestprädizierbarem Kriterium der Therapiegüte trägt "BDI Score Reduktion" etwas mehr bei als "Glucocorticoid Reduktion" bei.

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Selbskontrollfragen

1. Definieren Sie den Begriff der Korrelation.
2. Definieren Sie den Begriff der Stichprobenkorrelation.
3. Geben Sie das Theorem zu Korrelation und linear-affiner Abhängigkeit wieder.
4. Geben Sie das Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen wieder.
5. Erläutern Sie das Anwendungsszenario einer Kanonischen Korrelationsanalyse.
6. Erläutern Sie das Ziel einer Kanonischen Korrelationsanalyse.
7. Erläutern Sie die Begriffe "bester Prädiktor" und "am besten prädizierbares Kriterium".
8. Geben Sie die Definition Kanonischer Koeffizientenvektoren, Variate und Korrelationen an.
9. Geben Sie das Theorem zu den Eigenschaften kanonischer Korrelationen und Variate wieder.
10. Geben Sie die Definition für Schätzer kanonischer Korrelationen und Koeffizientenvektoren wieder.
11. Skizzieren Sie die Durchführung einer kanonischen Korrelationsanalyse.

Referenzen

Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. *Multivariate Analysis*. Probability and Mathematical Statistics. London ; New York: Academic Press.

Uurtio, Viivi, João M. Monteiro, Jaz Kandola, John Shawe-Taylor, Delmiro Fernandez-Reyes, and Juho Rousu. 2018. "A Tutorial on Canonical Correlation Methods." *ACM Computing Surveys* 50 (6): 1–33. https://doi.org/10.1145/3136624.