Multivariate Verfahren

MSc Psychologie & MSc Klinische Psychologie und Psychotherapie
Wintersemester 2024/2025

Joram Soch

(7) Kanonische Korrelationsanalyse

Datenanalyseszenarien

UV	AV	Datenanalysemethoden
univariat	univariat	Korrelation, einfache lineare Regression, T-Tests, ANOVA
multivariat	univariat	multiple Korrelation, multiple Regression, Allgemeines Lineares Modell
univariat	multivariat	T^2-Tests, einfaktorielle multivariate Varianzanalyse (MANOVA)
multivariat	multivariat	Kanonische Korrelation, multivariates Allgemeines Lineares Modell

Korrelation, einfache lineare Regression, T-Tests, ANOVA

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}

multiple Korrelation, multiple Regression, Allgemeines Lineares Modell

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}

T^2-Tests, einfaktorielle multivariate Varianzanalyse (MANOVA)

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}

Kanonische Korrelationsanalyse, multivariates Allgemeines Lineares Modell

UV			AV		
x_1	$\cdots$	x_{m_x}	y_1	$\cdots$	y_{m_y}
x_{11}	$\cdots$	x_{m_x1}	y_{11}	$\cdots$	y_{m_y1}
x_{12}	$\cdots$	x_{m_x2}	y_{12}	$\cdots$	y_{m_y2}
x_{13}	$\cdots$	x_{m_x3}	y_{13}	$\cdots$	y_{m_y3}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{1n}	$\cdots$	x_{m_xn}	y_{1n}	$\cdots$	y_{m_yn}

Datenanalyseszenarien

UV	AV	Datenanalysemethoden
univariat	univariat	Korrelation, einfache lineare Regression, T-Tests, ANOVA
multivariat	univariat	multiple Korrelation, multiple Regression, Allgemeines Lineares Modell
univariat	multivariat	T^2-Tests, einfaktorielle multivariate Varianzanalyse (MANOVA)
multivariat	multivariat	Kanonische Korrelation, multivariates Allgemeines Lineares Modell

Korrelation

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Beweise

Korrelation

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Beweise

Anwendungsszenario

Psychotherapie

Mehr Therapiestunden

$\Rightarrow$ Höhere Wirksamkeit?

Unabhängige Variable

- Anzahl Therapiestunden

Abhängige Variable

- BDI Score Reduktion

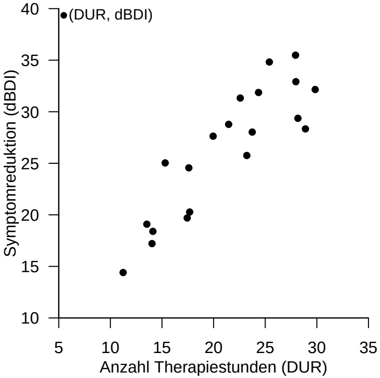
Beispieldatensatz

$i = 1, ..., 20$ Patient:innen, `dBDI` Symptomreduktion bei Patient:in i, `DUR` Anzahl Therapiestunden von Patient:in i

DUR	dBDI
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

Korrelation

Beispieldatensatz



$\Rightarrow$ Wie stark hängen Anzahl Therapiestunden und Symptomreduktion zusammen?

Definition (Korrelation)

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

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

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

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.
- Wir haben bereits gesehen, dass aus der Unabhängigkeit von ξ und υ, folgt dass $\rho(\xi, \upsilon) = 0$.
- Wenn $\rho(\xi, \upsilon) = 0$ ist, sind aber ξ und υ nicht notwendigerweise unabhängig voneinander.

Definition (Stichprobenkorrelation)

$\{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathbb{R}^2$ sei ein Datensatz. 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

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

- 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 = "Kanonische_Korrelationsanalyse.csv"                                  # Dateiname
D     = read.table(fname, sep = ",", header = TRUE)                           # Laden als Datenframe
x_i   = D$DUR                                                                 # x_i Werte
y_i   = D$dBDI                                                                # 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] 0.8829308
```

```r
# automatische Berechnung mit der R-Funktion cor()
r_xy  = cor(x_i,y_i)                                                          # r_{xy}
print(r_xy)                                                                   # Ausgabe
```

```
[1] 0.8829308
```

$\Rightarrow$ Anzahl Therapiestunden und Symptomreduktion sind hochkorreliert.

Mechanik der Kovariationsterme

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

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

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

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

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

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

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

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

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

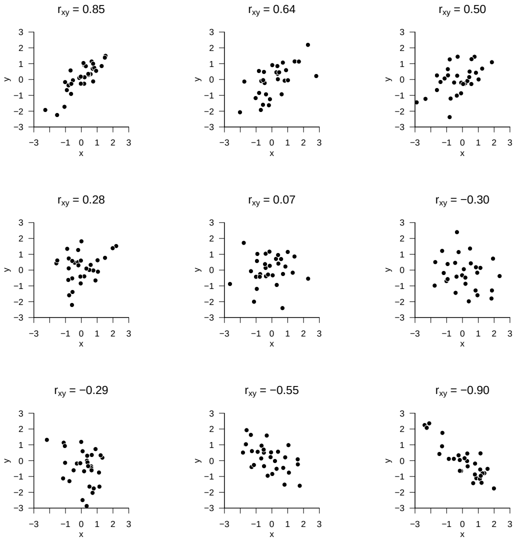
häufige richtungsgleiche Abweichung der x_i und y_i von ihren Mittelwerten $\Rightarrow$ positive Korrelation

häufige richtungsungleiche Abweichung der x_i und y_i von ihren Mittelwerten $\Rightarrow$ negative Korrelation

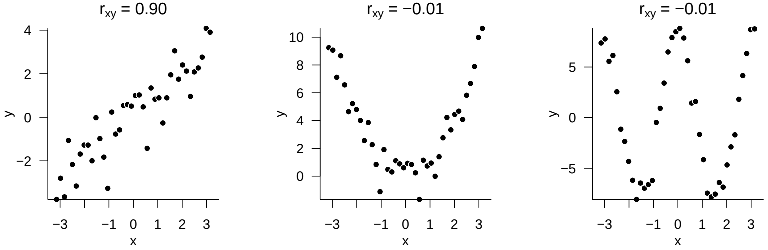
keine häufigen richtungsgleichen oder richtungsungleichen Abweichungen $\Rightarrow$ keine Korrelation

Korrelation

Beispiele



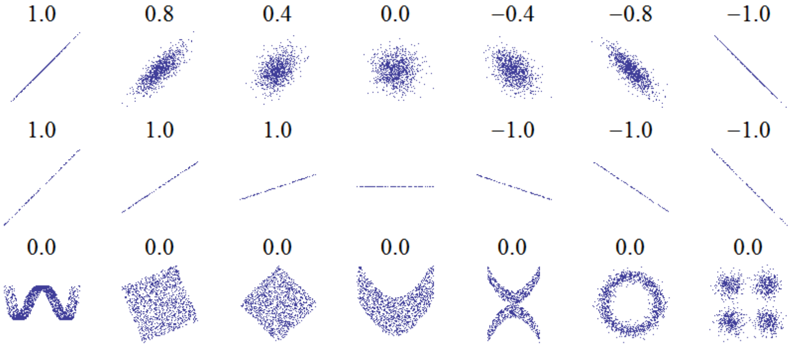
Funktionale Abhängigkeiten und Stichprobenkorrelation



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

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

Weitere Beispiele zur Stichprobenkorrelation



(Quelle: *Wikimedia Commons*: "Correlation_examples.png"; Lizenz: gemeinfrei.)

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

$$\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta) = \alpha\gamma\mathbb{C}(\xi, \upsilon) \tag{6}$$

und

$$\rho(\alpha\xi + \beta, \gamma\upsilon + \delta) = \rho(\xi, \upsilon). \tag{7}$$

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

$$
\begin{aligned}
\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta) &= \mathbb{E}((\alpha\xi + \beta - \mathbb{E}(\alpha\xi + \beta))(\gamma\upsilon + \delta - \mathbb{E}(\gamma\upsilon + \delta))) \\
&= \mathbb{E}((\alpha\xi + \beta - \alpha\mathbb{E}(\xi) - \beta)(\gamma\upsilon + \delta - \gamma\mathbb{E}(\gamma\upsilon) - \delta)) \\
&= \mathbb{E}(\alpha(\xi - \mathbb{E}(\xi))(\gamma(\upsilon - \mathbb{E}(\upsilon)))) \\
&= \mathbb{E}(\alpha\gamma((\xi - \mathbb{E}(\xi))(\upsilon - \mathbb{E}(\upsilon)))) \\
&= \alpha\gamma\mathbb{C}(\xi, \upsilon)\ .
\end{aligned}
\tag{8}
$$

Also folgt

$$
\begin{aligned}
\rho(\alpha\xi + \beta, \gamma\upsilon + \delta) &= \frac{\mathbb{C}(\alpha\xi + \beta, \gamma\upsilon + \delta)}{\sqrt{\mathbb{V}(\alpha\xi + \beta)}\sqrt{\mathbb{V}(\gamma\upsilon + \delta)}} \\
&= \frac{\alpha\gamma\mathbb{C}(\xi, \upsilon)}{\sqrt{\alpha^2\mathbb{V}(\xi)}\sqrt{\gamma^2\mathbb{V}(\upsilon)}} \\
&= \frac{\alpha\gamma\mathbb{C}(\xi, \upsilon)}{\alpha\mathbb{S}(\xi)\gamma\mathbb{S}(\upsilon)} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\mathbb{S}(\xi)\mathbb{S}(\upsilon)} \\
&= \rho(\xi, \upsilon)\ .
\end{aligned}
\tag{9}
$$

□

Anwendungsszenario zur Kanonischen Korrelationsanalyse

Therapiegüte als Therapieerfolgsfaktor?

Unabhängige Variablen

- Anzahl Therapiestunden
- Erfahrung Therapeut:in

⇒ Maß für Therapiegüte

Abhängige Variablen

- BDI Score Reduktion
- Glucocorticoid Reduktion

⇒ Maß für Therapieerfolg

Korrelation

Beispieldatensatz zur Kanonischen Korrelationsanalyse

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

`DUR` (x_1) Therapiedauer, `EXP` (x_2) Erfahrung Psychotherapeut:in

`dBDI` (y_1) BDI-Score-Reduktion, `dGLU` (y_2) Glukokortikoidplasmalevel-Reduktion

DUR	EXP	dBDI	dGLU
27.9	7.8	35.5	6.1
15.3	9.3	25.0	4.0
17.4	2.1	19.7	1.7
21.5	6.5	28.8	2.6
28.2	1.3	29.4	1.9
14.0	2.7	17.2	0.9
28.0	3.9	32.9	2.0
28.9	0.1	28.3	4.1
23.2	3.8	25.8	3.9
22.6	8.7	31.3	3.8
11.2	3.4	14.4	2.1
14.1	4.8	18.4	2.0
13.5	6.0	19.1	5.0
23.7	4.9	28.0	2.6
17.7	1.9	20.3	2.1
25.4	8.3	34.8	4.4
20.0	6.7	27.6	4.0
24.4	7.9	31.9	3.9
29.8	1.1	32.2	1.0
17.6	7.2	24.6	1.9

Korrelation

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Beweise

Modellformulierung

Grundzüge der Kanonischen Korrelationsanalyse

Zufallsvektoren

u.i.v. Realisierungen

x			y		
x_{11}	$\cdots$	$x_{m_x 1}$	y_{11}	$\cdots$	$y_{m_y 1}$
x_{12}	$\cdots$	$x_{m_x 2}$	y_{12}	$\cdots$	$y_{m_y 2}$
$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
x_{1n}	$\cdots$	$x_{m_x n}$	y_{1n}	$\cdots$	$y_{m_y n}$

$a \in \mathbb{R}^{m_x}, b \in \mathbb{R}^{m_y}$

Linearkombinationen von Zufallsvektorkomponenten

$a^T x$	$b^T y$
$\sum_{i=1}^{m_x} a_i x_{1i}$	$\sum_{i=1}^{m_y} b_i y_{1i}$
$\sum_{i=1}^{m_x} a_i x_{2i}$	$\sum_{i=1}^{m_y} b_i y_{2i}$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$\sum_{i=1}^{m_x} a_i x_{ni}$	$\sum_{i=1}^{m_y} b_i y_{ni}$

$\xi := a^T x, \upsilon = b^T y$

Korrelation von Zufallsvariablen

ξ	υ
ξ_1	υ_1
ξ_2	υ_2
$\vdots$	$\vdots$
$\vdots$	$\vdots$
ξ_n	υ_n

$$\max_{a,b} \rho(\xi, \upsilon) \text{ u.d.N. } \mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$$

Grundzüge der Kanonischen Korrelationsanalyse

Man betrachtet multivariate unabhängige Variablen und multivariate abhängige Variablen. Die unabhängigen Variablen werden dabei auch *Prädiktoren*, die abhängigen Variablen auch *Kriterien* genannt.

Beobachtete Werte der Prädiktoren werden als u.i.v. Realisierungen eines m_x-dimensionalen Zufallsvektors x interpretiert, beobachtete Werte der Kriterien werden als u.i.v. Realisierungen eines m_y-dimensionalen Zufallsvektors y interpretiert.

Man fragt nach der maximal möglichen Korrelation von Linearkombinationen von x und y. Wir bezeichnen dazu Linearkombinationen von x und y mit Vektoren $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ mit

$$\xi = a^{\mathrm{T}} x = a_1 x_1 + \cdots + a_{m_x} x_{m_x} \quad \text{und} \quad \upsilon = b^{\mathrm{T}} y = b_1 y_1 + \cdots + b_{m_y} y_{m_y}. \tag{10}$$

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

Die spezifische Linearkombinationen $\xi = a^{\mathrm{T}} x$ und $\upsilon = b^{\mathrm{T}} y$ für die $\rho(\xi, \upsilon)$ maximal ist, können dann als "bester Prädiktor" und als "am besten prädizierbares Kriterium" interpretiert werden. Um diese zu bestimmenm, fragt die Kanonische Korrelationsanalyse also 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 $\rho\left(a^{\mathrm{T}} x, b^{\mathrm{T}} y\right)$ und $\rho\left(\left(\alpha a^{\mathrm{T}}\right) x, \left(\beta b^{\mathrm{T}}\right) y\right)$ allerdings, wie im Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen gesehen, identisch. Man sucht deshalb Parameter $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$, für die $\rho(\xi, \upsilon) = \rho\left(a^{\mathrm{T}} x, b^{\mathrm{T}} y\right)$ maximal ist und für die $a^{\mathrm{T}} x$ und $b^{\mathrm{T}} y$ jeweils eine Varianz von 1 haben, also $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ gilt.

Modellformulierung

Grundzüge der Kanonischen Korrelationsanalyse

Da mit dem Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen die Varianzen zu verschiedenen skalaren Vielfachen von $a^{\mathrm{T}}x$ und $b^{\mathrm{T}}y$ verschieden sind, legt $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ die $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$, für die $\rho(\xi, \upsilon)$ maximal ist, eindeutig fest. Zur Bestimmung von $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ ist man also auf ein restringiertes Optimierungsproblem geführt.

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

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^m \tag{11}$$

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) 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, die im Folgenden diskutiert werden, auf diese Matrixprodukte angewendet werden. Generell wird in der Entwicklung nach Mardia, Kent, and Bibby (1979) ein restringierter Optimierungsansatz mithilfe der Lagrangefunktion zugunsten der Eigenanalyse von Matrixprodukten verworfen. Für die Entwicklung mit einem Lagrangeansatz, siehe zum Beispiel Anderson (2003) und die Originalarbeiten von H. Hotelling (1935) und Harold Hotelling (1936).

Theorem (Kovarianzmatrizen von Zufallsvektoren)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mit} \quad \mathbb{E}(z) := 0_m \tag{12}$$

ein m-dimensionaler Zufallsvektor bzw. sein Erwartungswertvektor. Dann kann die $m \times m$ Kovarianzmatrix z geschrieben werden als

$$\mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m}\,, \tag{13}$$

wobei

$$\begin{aligned}
\Sigma_{xx} &:= \mathbb{E}\left(xx^{\mathrm{T}}\right) \in \mathbb{R}^{m_x \times m_x} \\
\Sigma_{xy} &:= \mathbb{E}\left(xy^{\mathrm{T}}\right) \in \mathbb{R}^{m_x \times m_y} \\
\Sigma_{yx} &:= \mathbb{E}\left(yx^{\mathrm{T}}\right) \in \mathbb{R}^{m_y \times m_x} \\
\Sigma_{yy} &:= \mathbb{E}\left(yy^{\mathrm{T}}\right) \in \mathbb{R}^{m_x \times m_y}\,.
\end{aligned} \tag{14}$$

Bemerkungen

- Wenn der Erwartungswertvektor eines zusammengesetzten Zufallsvektors der Nullvektor ist, ergibt sich die Kovarianzmatrix des Zufallsvektors aus Erwartungswerten von äußeren Produkten der Bestandteile des Zufallsvektors.

Beweis

Nach Definition der Kovarianzmatrix eines Zufallsvektors gilt

$$
\begin{aligned}
\mathbb{C}(z) &= \mathbb{E}\left((z-\mathbb{E}(z))(z-\mathbb{E}(z))^{\mathrm{T}}\right) \\
&= \mathbb{E}\left((z-0_m)(z-0_m)^{\mathrm{T}}\right) \\
&= \mathbb{E}\left(zz^{\mathrm{T}}\right) \\
&= \mathbb{E}\left(\begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} x^{\mathrm{T}} & y^{\mathrm{T}} \end{pmatrix}\right) \\
&= \mathbb{E}\left(\begin{pmatrix} xx^{\mathrm{T}} & xy^{\mathrm{T}} \\ yx^{\mathrm{T}} & yy^{\mathrm{T}} \end{pmatrix}\right) \\
&= \begin{pmatrix} \mathbb{E}\left(xx^{\mathrm{T}}\right) & \mathbb{E}\left(xy^{\mathrm{T}}\right) \\ \mathbb{E}\left(yx^{\mathrm{T}}\right) & \mathbb{E}\left(yy^{\mathrm{T}}\right) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} .
\end{aligned}
\tag{15}
$$

$\square$

Theorem (Linearkombinationen von Zufallsvektorpartitionen)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mit} \quad \mathbb{E}(z) = 0_m \quad \text{und} \quad \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{16}$$

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

$$\xi := a^{\mathrm{T}} x \quad \text{und} \quad \upsilon := b^{\mathrm{T}} y \tag{17}$$

als Linearkombinationen der Komponenten von x und y definiert. Dann gilt:

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

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

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

Bemerkungen

- Die Varianz der Zufallsvariable $a^{\mathrm{T}} x$ ergibt sich als "quadrierte Linearkombination" von Σ_{xx}.
- Die Varianz der Zufallsvariable $b^{\mathrm{T}} y$ ergibt sich als "quadrierte Linearkombination" von Σ_{yy}.
- Die Korrelation der Zufallsvariablen $a^{\mathrm{T}} x$ und $b^{\mathrm{T}} y$ ergibt sich "quadrierte Linearkombination" von Σ_{xy}.

<u>Beweis</u>

(1) Wir betrachten zunächst die Varianz von ξ. Mit dem Varianzverschiebungssatz gilt:

$$
\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left(\xi\xi\right) - \mathbb{E}(\xi)\mathbb{E}(\xi) \\
&= \mathbb{E}\left((a^{\mathrm{T}}x)(a^{\mathrm{T}}x)\right) - \mathbb{E}\left(a^{\mathrm{T}}x\right)\mathbb{E}\left(a^{\mathrm{T}}x\right) \\
&= \mathbb{E}\left((a^{\mathrm{T}}x)(a^{\mathrm{T}}x)^{\mathrm{T}}\right) - \mathbb{E}\left(a^{\mathrm{T}}x\right)\mathbb{E}\left(a^{\mathrm{T}}x\right) \\
&= \mathbb{E}\left(a^{\mathrm{T}}xx^{\mathrm{T}}a\right) - \mathbb{E}\left(a^{\mathrm{T}}x\right)\mathbb{E}\left(a^{\mathrm{T}}x\right) \\
&= a^{\mathrm{T}}\mathbb{E}\left(xx^{\mathrm{T}}\right)a - a^{\mathrm{T}}\mathbb{E}(x)a^{\mathrm{T}}\mathbb{E}(x) \\
&= a^{\mathrm{T}}\mathbb{E}\left(xx^{\mathrm{T}}\right)a - a^{\mathrm{T}}0_{m_x}a^{\mathrm{T}}0_{m_x} \\
&= a^{\mathrm{T}}\Sigma_{xx}a\ .
\end{aligned}
\tag{18}
$$

(2) Der Beweis zur Varianz von υ folgt dann analog.

Beweis

(3) Mit der Definition der Zufallsvariablen sowie $\mathbb{V}(\xi) = \mathbb{V}(\upsilon) = 1$ und dem Kovarianzverschiebungssatz gilt:

$$
\begin{aligned}
\rho(\xi, \upsilon) &= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(\upsilon)}} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{1}\sqrt{1}} \\
&= \mathbb{C}(\xi, \upsilon) \\
&= \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) \\
&= \mathbb{E}\left((a^{\mathrm{T}}x)(b^{\mathrm{T}}y)\right) - \mathbb{E}(a^{\mathrm{T}}x)\mathbb{E}(b^{\mathrm{T}}y) \\
&= \mathbb{E}\left((a^{\mathrm{T}}x)(b^{\mathrm{T}}y)^{\mathrm{T}}\right) - \mathbb{E}(a^{\mathrm{T}}x)\mathbb{E}(b^{\mathrm{T}}y) \\
&= \mathbb{E}\left(a^{\mathrm{T}}xy^{\mathrm{T}}b\right) - \mathbb{E}(a^{\mathrm{T}}x)\mathbb{E}(b^{\mathrm{T}}y) \\
&= a^{\mathrm{T}}\mathbb{E}\left(xy^{\mathrm{T}}\right)b - a^{\mathrm{T}}\mathbb{E}(x)b^{\mathrm{T}}\mathbb{E}(y) \\
&= a^{\mathrm{T}}\mathbb{E}\left(xy^{\mathrm{T}}\right)b - a^{\mathrm{T}}0_{m_x}b^{\mathrm{T}}0_{m_y} \\
&= a^{\mathrm{T}}\Sigma_{xy}b\,.
\end{aligned}
\tag{19}
$$

$\square$

Definition (Kanonische Koeffizientenvektoren, Variate, Korrelationen)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mit} \quad \mathbb{E}(z) = 0_m \quad \text{und} \quad \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{20}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswertvektor und seine Kovarianzmatrix. Weiterhin sei

$$K := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{21}$$

mit der Singulärwertzerlegung

$$K = A \Lambda B^{\mathrm{T}}, \tag{22}$$

wobei

$$A := \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \end{pmatrix} \in \mathbb{R}^{m_x \times m_y} \quad \text{und} \quad B := \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix} \in \mathbb{R}^{m_y \times m_y} \tag{23}$$

die orthogonalen Matrix der Eigenvektoren von KK^{T} bzw. die orthogonale Matrix der Eigenvektoren von $K^{\mathrm{T}}K$ bezeichnen und

$$\Lambda := \mathrm{diag}\left(\sqrt{\lambda_1}, ..., \sqrt{\lambda_k}\right) \in \mathbb{R}^{m_y \times m_y}, \tag{24}$$

die Diagonalmatrix der Quadratwurzeln der zugehörigen absteigend geordneten Eigenwerte bezeichnet. Schließlich seien für $i = 1, ..., k := \min\{m_x, m_y\}$

$$a_i := \Sigma_{xx}^{-1/2} \alpha_i \in \mathbb{R}^{m_x} \quad \text{und} \quad b_i := \Sigma_{yy}^{-1/2} \beta_i \in \mathbb{R}^{m_y}. \tag{25}$$

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 $\xi_i := a_i^{\mathrm{T}} x$ und $\upsilon_i := b_i^{\mathrm{T}} y$ die $i$ten i*ten kanonischen Variaten* und

(3) $\rho_i := \sqrt{\lambda_i}$ die i*te kanonische Korrelation*.

Theorem (Eigenschaften kanonischer Korrelationen und Variaten)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mit} \quad \mathbb{E}(z) = 0_m \quad \text{und} \quad \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{26}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswertvektor und seine Kovarianzmatrix. Weiterhin seien für $i = 1, ..., k := \min\{m_x, m_y\}$ die kanonischen Koeffizientenvektoren a_i, b_i, die kanonischen Variaten ξ_i, 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

$$\phi_r = \max_{a,b} a^{\mathrm{T}} \Sigma_{xy} b \tag{27}$$

unter den Nebenbedingungen

$$a^{\mathrm{T}} \Sigma_{xx} a = 1, \quad b^{\mathrm{T}} \Sigma_{yy} b = 1, \quad a_i^{\mathrm{T}} \Sigma_{xx} a = 0 \quad \text{für} \quad i = 1, ..., r-1 \tag{28}$$

(1) den Wert $\phi_r = \rho_r$ hat und (2) dass ϕ_r diesen Wert bei $a = a_r$ und $b = b_r$ annimmt.

Bemerkungen

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

Modellformulierung

Simulationsbeispiel

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

$$p(x) = N(x; 0_4, I_4) \quad \text{und} \quad p(y|x) = N(y; Lx, G) \tag{29}$$

mit

$$L := \begin{pmatrix} 0.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 \end{pmatrix} \quad \text{und} \quad G := \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix}. \tag{30}$$

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

$$\begin{aligned} y_1 &= x_3 + \varepsilon_1 \\ y_2 &= x_1 + \varepsilon_2 \\ y_3 &= -x_4 + \varepsilon_3 \end{aligned} \tag{31}$$

mit

$$x_1 \sim N(0,1),\ x_3 \sim N(0,1),\ x_4 \sim N(0,1) \tag{32}$$

und

$$\varepsilon_1 \sim N(0, 0.2),\ \varepsilon_2 \sim N(0, 0.4),\ \varepsilon_3 \sim N(0, 0.3). \tag{33}$$

Simulationsbeispiel

Mit dem Theorem zu gemeinsamen Normalverteilungen (vgl. Einheit (4) Multivariate Normalverteilungen) ergibt sich, dass

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N(0_7, \Sigma) \tag{34}$$

mit

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{35}$$

wobei

$$\Sigma_{xx} = I_4, \quad \Sigma_{xy} = L^{\mathrm{T}}, \quad \Sigma_{yx} = L \quad \text{und} \quad \Sigma_{yy} = G + LL^{\mathrm{T}}. \tag{36}$$

Explizit ergibt sich also

$$\Sigma = \begin{pmatrix} I_4 & L^{\mathrm{T}} \\ L & G + LL^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.2 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & 1.3 \end{pmatrix} \tag{37}$$

Modellformulierung

Simulationsbeispiel

```
# R-Pakete für Matrizenrechnung
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
```

Simulationsbeispiel

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

Die kanonische Korrelationen und kanonischen Koeffizientenvektoren ergeben sich zu

```
rho_1 = 0.9129 , a_1^T = ( 0 0 -1 0 ), b_1^T = ( -0.9129 0 0 )
rho_2 = 0.8771 , a_2^T = ( 0 0 0 1 ),  b_2^T = ( 0 0 -0.8771 )
rho_3 = 0.8452 , a_3^T = ( -1 0 0 0 ), b_3^T = ( 0 -0.8452 0 )
```

Korrelation

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Beweise

Definition (Schätzer der kanonischen Korrelationsanalyse)

Für $i = 1, ..., n$ seien

$$z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \text{mit} \quad \mathbb{E}(z_i) := 0_m \quad \text{und} \quad \mathbb{C}(z_i) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{38}$$

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

$$C := \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{39}$$

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

$$\hat{a}_i := C_{xx}^{-1/2} \hat{\alpha}_i \in \mathbb{R}^{m_x}, \quad \hat{b}_i := C_{yy}^{-1/2} \hat{\beta}_i \in \mathbb{R}^{m_y} \quad \text{und} \quad \hat{\rho}_i := \sqrt{\hat{\lambda}_i} \tag{40}$$

Schätzer der $i$ten kanonischen Koeffizientenvektoren bzw. kanonischen Korrelationen. Dabei sind mit

$$\hat{K} := C_{xx}^{-1/2} C_{xy} C_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{41}$$

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

Bemerkungen

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

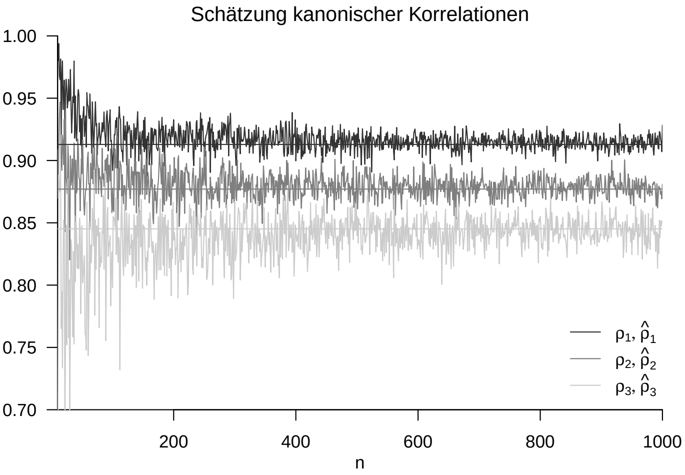
Simulationsbeispiel

```
# R-Pakete
library(MASS)
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(solve(Sigma_xx)) %*% Sigma_xy %*% sqrtm(solve(Sigma_yy))
ALB      = svd(K)
A        = ALB$u
Lambda   = ALB$d
B        = ALB$v
rho      = Lambda
a        = sqrtm(solve(Sigma_xx)) %*% A
b        = sqrtm(solve(Sigma_yy)) %*% B
```

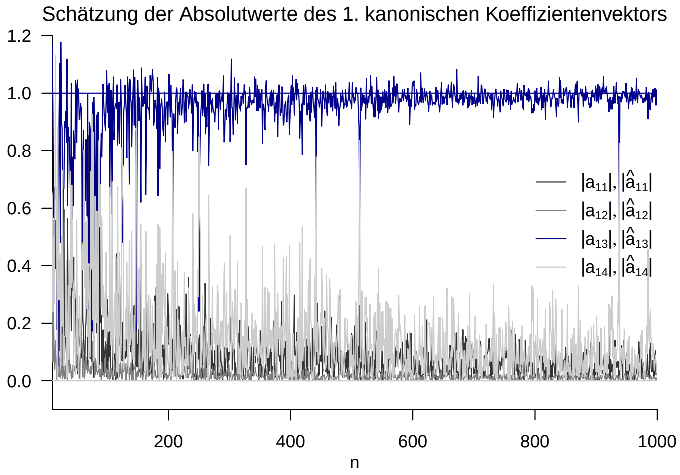
Simulationsbeispiel

```
# Vorbereitung
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)

# Simulationen
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])

    # Stichprobenkovarianzmatrix
    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(solve(C_xx)) %*% C_xy %*% sqrtm(solve(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(solve(C_xx)) %*% A_hat
    b_hat       = sqrtm(solve(C_yy)) %*% B_hat
    rho_hat[,i] = as.matrix(Lambda_hat)
    a_1_hat[,i] = a_hat[,1]
}
```

Simulationsbeispiel



Simulationsbeispiel



Anwendungsbeispiel

Therapiegüte als Therapieerfolgsfaktor?

Unabhängige Variablen

- Anzahl Therapiestunden
- Erfahrung Therapeut:in

⇒ Maß für Therapiegüte

Abhängige Variablen

- BDI Score Reduktion
- Glucocorticoid Reduktion

⇒ Maß für Therapieerfolg

Modelschätzung

Beispieldatensatz zur Kanonischen Korrelationsanalyse

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

`DUR` (x_1) Therapiedauer, `EXP` (x_2) Erfahrung Psychotherapeut:in

`dBDI` (y_1) BDI-Score-Reduktion, `dGLU` (y_2) Glukokortikoidplasmalevel-Reduktion

DUR	EXP	dBDI	dGLU
27.9	7.8	35.5	6.1
15.3	9.3	25.0	4.0
17.4	2.1	19.7	1.7
21.5	6.5	28.8	2.6
28.2	1.3	29.4	1.9
14.0	2.7	17.2	0.9
28.0	3.9	32.9	2.0
28.9	0.1	28.3	4.1
23.2	3.8	25.8	3.9
22.6	8.7	31.3	3.8
11.2	3.4	14.4	2.1
14.1	4.8	18.4	2.0
13.5	6.0	19.1	5.0
23.7	4.9	28.0	2.6
17.7	1.9	20.3	2.1
25.4	8.3	34.8	4.4
20.0	6.7	27.6	4.0
24.4	7.9	31.9	3.9
29.8	1.1	32.2	1.0
17.6	7.2	24.6	1.9

Anwendungsbeispiel

```r
# R-Pakete
library(expm)

# Datenpräprozessierung
D          = read.csv("Kanonische_Korrelationsanalyse.csv")
x          = as.matrix(cbind(D$DUR,  D$EXP))
y          = as.matrix(cbind(D$dBDI, D$dGLU))
n          = nrow(x)
m_x        = ncol(x)
m_y        = ncol(y)
Y          = t(cbind(x,y))

# Stichprobenkovarianzmatrix
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(solve(C_xx)) %*% C_xy %*% sqrtm(solve(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(solve(C_xx)) %*% A_hat
b_hat      = sqrtm(solve(C_yy)) %*% B_hat
rho_hat    = as.matrix(Lambda_hat)
```

```
rho_hat_1 :  0.9950575
a_hat_1   :  -0.1623409 -0.173979
b_hat_1   :  -0.1554175 -0.05025419
rho_hat_2 :  0.5010358
a_hat_2   :  -0.06026274 0.3118808
b_hat_2   :  -0.08128072 0.7773036
```

Anwendungsbeispiel

Kanonische Korrelationsanalyse mit der R-Funktion cancor():

```
# Datenpräprozessierung
D           = read.csv("Kanonische_Korrelationsanalyse.csv")
x           = as.matrix(cbind(D$DUR , D$EXP))
y           = as.matrix(cbind(D$dBDI, D$dGLU))
cca         = cancor(x,y)
```

```
# Ausgabe
cat(  "rho_hat_1 : ", cca$cor[1],
    "\nrho_hat_2 : ", cca$cor[2], "\n")
```

```
rho_hat_1 :  0.9950575
rho_hat_2 :  0.5010358
```

Modellschätzung

Anwendungsbeispiel

Die geschätzte maximale Korrelation einer Linearkombination von (`DUR`, `EXP`) und (`dBDI`, `dGLU`) ist 0.99.

- (`DUR`, `EXP`) und (`dBDI`, `dGLU`) sind multivariat also hochgradig korreliert.

Basierend auf der Schätzung der kanonischen Koeffizientenvektoren ergibt sich

- $\xi = 0.16$ DUR $+ 0.17$ EXP als "bester Prädiktor"
- $\upsilon = 0.15$ dBDI $+ 0.05$ dGLU als "am besten prädizierbares Kriterium"

Therapiedauer `DUR` und Therapeut:innenerfahrung `EXP` scheinen zur bestmöglichen Prädiktion der Therapiegüte also in etwa gleichbedeutend, bei dem bestprädizierbarem Kriterium der Therapiegüte trägt die BDI-Score-Reduktion `dBDI` etwas mehr bei als die Glukokortikoidplasmalevel-Reduktion `dGLU` – alles angesichts der betrachteten Datenskalierung.

Korrelation

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Beweise

Selbstkontrollfragen

1. Geben Sie die Definition der Korrelation wieder.
2. Geben Sie die Definition der Stichprobenkorrelation wieder.
3. Erläutern Sie, warum die Stichprobenkorrelation im Falle nicht-linearer Zusammenhänge von abhängigen und unabhängigen Variablen zu irreführenden Ergebnissen führen kann.
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 wieder.
9. Geben Sie das Theorem zu den Eigenschaften kanonischer Korrelationen und Variaten wieder.
10. Geben Sie die Definition für Schätzer kanonischer Korrelationen und Koeffizientenvektoren wieder.
11. Erläutern Sie die Durchführung einer kanonischen Korrelationsanalyse für einen Datensatz.

Korrelation

Modellformulierung

Modellschätzung

Selbstkontrollfragen

Beweise

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 mit der orthonormalen Matrix $Q \in \mathbb{R}^{m\times m}$ der Eigenvektoren von A und der Diagonalmatrix $\Lambda = \text{diag}(\lambda_i) \in \mathbb{R}^{m\times m}$ der zugehörigen Eigenwerte $\lambda_1, ..., \lambda_m$ von A definiert als

$$A^{r/s} = Q\Lambda^{r/s}Q^{\mathrm{T}} \quad \text{mit} \quad \Lambda^{r/s} = \text{diag}\left(\lambda_i^{r/s}\right). \tag{42}$$

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

$$A^{1/2} = Q\Lambda^{1/2}Q^{\mathrm{T}} \quad \text{mit} \quad \Lambda^{1/2} = \text{diag}\left(\lambda_i^{1/2}\right). \tag{43}$$

Bemerkungen

- Offenbar gilt
$$\left(A^{1/2}\right)^2 = Q\Lambda^{1/2}Q^{\mathrm{T}}Q\Lambda^{1/2}Q^{\mathrm{T}} = Q\Lambda^{1/2}\Lambda^{1/2}Q^{\mathrm{T}} = Q\Lambda Q^{\mathrm{T}} = A. \tag{44}$$
- Weiterhin gilt
$$\left(A^{-1/2}\right)^2 = Q\Lambda^{-1/2}Q^{\mathrm{T}}Q\Lambda^{-1/2}Q^{\mathrm{T}} = Q\Lambda^{-1/2}\Lambda^{-1/2}Q^{\mathrm{T}} = Q\Lambda^{-1}Q^{\mathrm{T}} = A^{-1}. \tag{45}$$
- Schließlich gilt
$$\begin{aligned} A^{-1/2}AA^{-1/2} &= Q\Lambda^{-1/2}Q^{\mathrm{T}}Q\Lambda Q^{\mathrm{T}}Q\Lambda^{-1/2}Q^{\mathrm{T}} \\ &= Q\Lambda^{-1/2}\Lambda\Lambda^{-1/2}Q^{\mathrm{T}} \\ &= Q\Lambda\Lambda^{-1}Q^{\mathrm{T}} \\ &= I_m \end{aligned} \tag{46}$$

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 zum Eigenwert λ ist.

Bemerkungen

- Für einen Beweis, siehe Mardia, Kent, and Bibby (1979), S. 468.
- Der folgende **R**-Code zeigt die Gültigkeit des Theorems für beispielhafte Matrizen.

```
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$values[1:2],
    "\nBAw mit w = Bv    :", B %*% A %*% w,
    "\n lw mit w = Bv    :", EBA$values[1] * w, "\n")
```

```
Eigenwerte von AB : 85.57934 0.4206623
Eigenwerte von BA : 85.57934 0.4206623
BAw mit w = Bv    : -191.1333 -416.7586 -642.3839
 lw mit w = Bv    : -191.1333 -416.7586 -642.3839
```

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 $Aab^{\mathrm{T}}B \in \mathbb{R}^{n\times n}$ gleich $b^{\mathrm{T}}BAa$ mit zugehörigem Eigenvektor Aa ist.

Bemerkungen

- Für einen Beweis siehe Mardia, Kent, and Bibby (1979), S. 468.
- Der folgende **R**-Code zeigt die Gültigkeit des Theorems für beispielhafte Matrizen.

```
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 A a b^T B \in \mathbb{R}^{4 x 4}
cat(  "Eigenwerte von A a b^T B :", EAabTB$values,
    "\nb^T B A a               :", t(b) %*% B %*% A %*% a,
    "\nA a                     :", A %*% a,
    "\n(A a b^T B) A a         :",(A %*% a %*% t(b) %*% B) %*% A %*% a,                  # = Mv
    "\n(b^T B A a) A a         :", as.vector((t(b) %*% B %*% A %*% a)) * (A %*% a), "\n")  # = \lambda v
```

```
Eigenwerte von A a b^T B : 2620 0
b^T B A a                : 2620
A a                      : 14 32
(A a b^T B) A a          : 36680 83840
(b^T B A a) A a          : 36680 83840
```

Theorem (Maximierung quadratischer Formen mit Nebenbedingungen)

$A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{m\times m}$ pd. 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

$$\max_x x^{\mathrm{T}} A x \quad \text{unter der Nebenbedingung} \quad x^{\mathrm{T}} B x = 1. \tag{47}$$

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

$$f : \mathbb{R}^m \to \mathbb{R}, x \mapsto f(x) := x^{\mathrm{T}} A x, \tag{48}$$

dass

$$v_1 = \arg\max_x x^{\mathrm{T}} A x \quad \text{unter der Nebenbedingung} \quad x^{\mathrm{T}} B x = 1 \tag{49}$$

und dass

$$\lambda_1 = \max_x x^{\mathrm{T}} A x \quad \text{unter der Nebenbedingung} \quad x^{\mathrm{T}} B x = 1. \tag{50}$$

Theorem (Eigenschaften kanonischer Korrelationen und Variaten)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mit} \quad \mathbb{E}(z) = 0_m \quad \text{und} \quad \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{51}$$

ein m-dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswertvektor und seine Kovarianzmatrix. 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

$$\phi_r = \max_{a,b} a^{\mathrm{T}} \Sigma_{xy} b \tag{52}$$

unter den Nebenbedingungen

$$a^{\mathrm{T}} \Sigma_{xx} a = 1, \quad b^{\mathrm{T}} \Sigma_{yy} b = 1, \quad a_i^{\mathrm{T}} \Sigma_{xx} a = 0 \quad \text{für} \quad i = 1, ..., r-1 \tag{53}$$

(1) den Wert $\phi_r = \rho_r$ hat und (2) dass ϕ_r diesen Wert bei $a = a_r$ und $b = b_r$ annimmt.

Bemerkungen

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

Beweise

Beweis

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

$$y := B^{1/2}x \quad \Leftrightarrow \quad x = B^{-1/2}y \tag{54}$$

Dann kann mit der symmetrischen Matrix

$$K := B^{-1/2}AB^{-1/2} \in \mathbb{R}^{m\times m} \tag{55}$$

das Optimierungsproblem (47) geschrieben werden als

$$\max_y y^{\mathrm{T}}Ky \quad \text{unter der Nebenbedingung} \quad y^{\mathrm{T}}y = 1. \tag{56}$$

Dies gilt, weil

$$\max_x x^{\mathrm{T}}Ax \quad \Leftrightarrow \quad \max_y \left(B^{-1/2}y\right)^{\mathrm{T}} A \left(B^{-1/2}y\right) \quad \Leftrightarrow \quad \max_y y^{\mathrm{T}}B^{-1/2}AB^{-1/2}y \quad \Leftrightarrow \quad \max_y y^{\mathrm{T}}Ky \tag{57}$$

und

$$x^{\mathrm{T}}Bx = 1 \quad \Leftrightarrow \quad y^{\mathrm{T}}B^{-1/2}BB^{-1/2}y = 1 \quad \Leftrightarrow \quad y^{\mathrm{T}}y = 1. \tag{58}$$

Weil K eine symmetrische Matrix ist, existiert die Orthonormalzerlegung (vgl. Einheit (3) Eigenanalyse)

$$K = Q\Lambda Q^{\mathrm{T}}, \tag{59}$$

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

Beweise

Beweis (fortgeführt)

Mit der orthogonalen Matrix Q aus obiger Orthornomalzerlegung sei nun

$$z := Q^{\mathrm{T}} y \quad \Leftrightarrow \quad y := Qz. \tag{60}$$

Dann kann das Optimierungsproblem (56) geschrieben werden als

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \quad \text{unter der Nebenbedingung} \quad z^{\mathrm{T}} z = 1, \tag{61}$$

weil

$$\max_y y^{\mathrm{T}} K y \quad \Leftrightarrow \quad \max_z (Qz)^{\mathrm{T}} K (Qz) \quad \Leftrightarrow \quad \max_z z^{\mathrm{T}} Q^{\mathrm{T}} Q \Lambda Q^{\mathrm{T}} Q z \quad \Leftrightarrow \quad \max_z z^{\mathrm{T}} \Lambda z \quad \Leftrightarrow \quad \max_z \sum_{i=1}^m \lambda_i z_i^2 \tag{62}$$

und

$$y^{\mathrm{T}} y = 1 \quad \Leftrightarrow \quad (Qz)^{\mathrm{T}} Q z = 1 \quad \Leftrightarrow \quad z^{\mathrm{T}} Q^{\mathrm{T}} Q z = 1 \quad \Leftrightarrow \quad z^{\mathrm{T}} z = 1. \tag{63}$$

Beweis (fortgeführt)

Die Eigenwerte von K bzw. Diagonalelemente von Λ seien nun absteigend sortiert, also $\lambda_1 \geq \cdots \geq \lambda_m$. Dann gilt für das Optimierungsproblem (61), dass

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_1, \tag{64}$$

weil

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \max_z \sum_{i=1}^m \lambda_1 z_i^2 = \lambda_1 \max_z \sum_{i=1}^m z_i^2 = \lambda_1, \tag{65}$$

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

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 = \lambda_1 \tag{66}$$

für $z := e_1 = (1, 0, ..., 0)^{\mathrm{T}}$. Zusammenfassend heißt das, dass $z = e_1$ eine Lösung des Optimierungsproblem (61) ist und dass λ_1 das entsprechende Maximum ist.

Beweis (fortgeführt)

Damit ergibt sich aber sofort, dass dann

$$y = Qz = Qe_1 = q_1 \quad \text{und} \quad x = B^{-1/2}q_1 \tag{67}$$

Lösungen der äquivalenten Optimierungsprobleme (56) bzw. (47) 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

$$B^{-1/2}B^{-1/2}A = B^{-1}A \tag{68}$$

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

$$\max_x x^{\mathrm{T}}Ax \quad \text{unter der Nebenbedingung} \quad x^{\mathrm{T}}Bx = 1 \tag{69}$$

ist. $\square$

Beweis (fortgeführt)

Wir betrachten das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a,b} \left(a^{\mathrm{T}} \Sigma_{xy} b\right)^2 \quad \text{u.d.N.} \quad a^{\mathrm{T}} \Sigma_{xx} a = 1, \quad b^{\mathrm{T}} \Sigma_{yy} b = 1, \quad a_i^{\mathrm{T}} \Sigma_{xx} a = 0,\ i = 1, ..., r-1. \qquad (70)$$

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

$$\phi_r^2 = \max_{a} \left(\max_{b} \left(a^{\mathrm{T}} \Sigma_{xy} b\right)^2 \quad \text{u.d.N.} \quad b^{\mathrm{T}} \Sigma_{yy} b = 1\right) \quad \text{u.d.N.} \quad a^{\mathrm{T}} \Sigma_{xx} a = 1, \quad a_i^{\mathrm{T}} \Sigma_{xy} a = 0,\ i = 1, ..., r-1 \qquad (71)$$

von innen nach außen.

Schritt (1)

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

$$\max_{b} \left(a^{\mathrm{T}} \Sigma_{xy} b\right)^2 \quad \text{u.d.N.} \quad b^{\mathrm{T}} \Sigma_{yy} b = 1. \qquad (72)$$

Dieses Optimierungsproblem kann geschrieben werden als

$$\max_{b} b^{\mathrm{T}} \Sigma_{yx} a a^{\mathrm{T}} \Sigma_{xy} b \quad \text{u.d.N.} \quad b^{\mathrm{T}} \Sigma_{yy} b = 1, \qquad (73)$$

weil gilt, dass

$$\left(a^{\mathrm{T}} \Sigma_{xy} b\right)^2 = \left(a^{\mathrm{T}} \Sigma_{xy} b\right)\left(a^{\mathrm{T}} \Sigma_{xy} b\right) = \left(a^{\mathrm{T}} \Sigma_{xy} b\right)^{\mathrm{T}} a^{\mathrm{T}} \Sigma_{xy} b = b^{\mathrm{T}} \Sigma_{yx} a a^{\mathrm{T}} \Sigma_{xy} b. \qquad (74)$$

Beweise

Beweis (fortgeführt)

Das Optimierungsproblem (73) kann nun mithilfe des Theorems zur Maximierung quadratischer Formen mit Nebenbedingen gelöst werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yx} a a^{\mathrm{T}} \Sigma_{xy} \quad \text{und} \quad B := \Sigma_{yy}. \tag{75}$$

Dann hat (73) die Form

$$\max_b b^{\mathrm{T}} A b \quad \text{unter der Nebenbedingung} \quad b^{\mathrm{T}} B b = 1. \tag{76}$$

Das Maximum von (76) entspricht nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen dem größten Eigenwert von

$$B^{-1} A = \Sigma_{yy}^{-1} \Sigma_{yx} a a^{\mathrm{T}} \Sigma_{xy}. \tag{77}$$

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

$$A := \Sigma_{yy}^{-1} \Sigma_{yx}, \quad b := a, \quad B := \Sigma_{xy} \tag{78}$$

und erhalten für den betreffenden Eigenwert

$$\lambda_a = b^{\mathrm{T}} B A a = a^{\mathrm{T}} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a. \tag{79}$$

als Lösung (Maximum) des restringierten Optimierungsproblems

$$\max_b \left(a^{\mathrm{T}} \Sigma_{xy} b \right)^2 \quad \text{unter der Nebenbedingung} \quad b^{\mathrm{T}} \Sigma_{yy} b = 1. \tag{80}$$

Beweis (fortgeführt)

Schritt (2)

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

$$\phi_r^2 = \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \quad \text{u.d.N.} \quad a^T \Sigma_{xx} a = 1, \quad a_i^T \Sigma_{xx} a = 0, i = 1, ..., r-1. \tag{81}$$

Dazu halten wir zunächst fest, dass (81) mit den Definitionen von α_i und K in der Definition der Kanonischen Koeffizientenvektoren, Variaten, und Korrelationen geschrieben werden kann als

$$\phi_r^2 = \max_\alpha \alpha^T K K^T \alpha \quad \text{u.d.N.} \quad \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0, i = 1, ..., r-1, \tag{82}$$

denn

$$\begin{aligned}
\phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \quad \text{u.d.N.} \quad a^T \Sigma_{xx} a = 1, \quad a_i^T \Sigma_{xx} a = 0 \\
&= \max_\alpha a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \quad \text{u.d.N.} \quad \alpha^T {\Sigma_{xx}^{-1/2}} \Sigma_{xx} {\Sigma_{xx}^{-1/2}} \alpha = 1, \quad \alpha_i^T {\Sigma_{xx}^{-1/2}} \Sigma_{xx} {\Sigma_{xx}^{-1/2}} \alpha = 0 \\
&= \max_\alpha \alpha^T {\Sigma_{xx}^{-1/2}} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} {\Sigma_{xx}^{-1/2}} \alpha \quad \text{u.d.N.} \quad \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0 \\
&= \max_\alpha \alpha^T {\Sigma_{xx}^{-1/2}} \Sigma_{xy} {\Sigma_{yy}^{-1/2}} {\Sigma_{yy}^{-1/2}} \Sigma_{yx} {\Sigma_{xx}^{-1/2}} \alpha \quad \text{u.d.N.} \quad \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0 \\
&= \max_\alpha \alpha^T K K^T \alpha \quad \text{u.d.N.} \quad \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0
\end{aligned} \tag{83}$$

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 (82) der größte Eigenwert von KK^{T} mit seinem assoziierten Eigenvektor. Die Nebenbedingung $\alpha_i^{\mathrm{T}}\alpha = 0$ schränkt diese Wahl auf den rt-größten Eigenwert und seinen assoziierten Eigenvektor α_r ein. Mit der Definition von Eigenwerten und Eigenvektoren gilt also

$$\phi_r^2 = \alpha_r^{\mathrm{T}} KK^{\mathrm{T}} \alpha_r = \alpha_r^{\mathrm{T}} \lambda_r \alpha_r = \lambda_r \alpha_r^{\mathrm{T}} \alpha_r = \lambda_r. \tag{84}$$

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 $a^{\mathrm{T}}\Sigma_{xy}b$ ergibt mit

$$K = A\Lambda B^{\mathrm{T}} \quad \Leftrightarrow \quad KB = A\Lambda B^{\mathrm{T}}B \quad \Leftrightarrow \quad KB = A\Lambda \quad \Leftrightarrow \quad K\beta_r = \alpha_r \lambda_r^{1/2}, \tag{85}$$

dass

$$a_r^{\mathrm{T}} \Sigma_{xy} b_r = \alpha_r^{\mathrm{T}} \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \beta_r = \alpha_r^{\mathrm{T}} K \beta_r = \alpha_r^{\mathrm{T}} \alpha_r \lambda_r^{1/2} = \rho_r. \tag{86}$$

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

Referenzen

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

Hotelling, H. 1935. "The Most Predictable Criterion." *Journal of Educational Psychology* 26 (2): 139–42. https://doi.org/10.1037/h0058165.

Hotelling, Harold. 1936. "Relations Between Two Sets of Variates." *Biometrika* 28 (3/4): 321. https://doi.org/10.2307/2333955.

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.