# 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 | $\mathrm{T}^2$-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 | | |
|---|---|---|---|
| $x_1$ | $y_1$ | $\cdots$ | $y_m$ |
| $x_{11}$ | $y_{12}$ | $\cdots$ | $y_{m1}$ |
| $x_{12}$ | $y_{13}$ | $\cdots$ | $y_{m2}$ |
| $x_{13}$ | $y_{14}$ | $\cdots$ | $y_{m3}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_{1n}$ | $y_{1n}$ | $\cdots$ | $y_{mn}$ |

$\text{T}^2$-Tests

Einfaktorielle MANOVA

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

**Kanonische Korrelationsanalyse**

Multivariates Allgemeines Lineares Modell

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

**Korrelation**

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

Modellformulierung

Modellschätzung

Selbstkontrollfragen

## Anwendungsszenario

Psychotherapie

Mehr Therapiestunden

$\Rightarrow$ Höhere Wirksamkeit?

Unabhängige Variable

- Anzahl Therapiestunden

Abhängige Variable

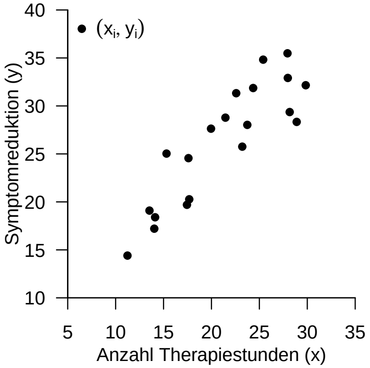
- BDI Score Reduktion

## 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?

# Korrelation

## Definition (Korrelation)

Die *Korrelation* zweier Zufallsvariablen $\xi$ und $\upsilon$ 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 $\xi$ und $\upsilon$ und $\mathbb{V}(\xi)$ und $\mathbb{V}(\upsilon)$ die Varianzen von $\xi$ und $\upsilon$ bezeichnen.

Für eine Einführung zur Korrelation siehe die entsprechenden BSc Lehreinheiten

- Erwartungswert, Varianz, Kovarianz
- Korrelation

Bemerkungen

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

## Definition (Stichprobenkorrelation)

$(x_1, y_1), ..., (x_n, y_n)$ seien unabhängigie Realisierungen eines Zufallsvektors $(\xi, \upsilon)$. 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

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

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

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

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

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

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

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

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

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

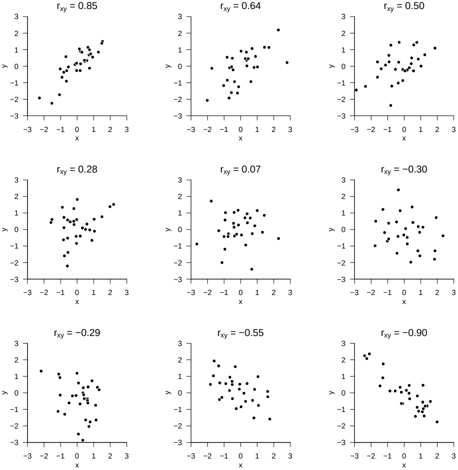
Häufige richtungsgleiche Abweichung der $x_i$ und $y_i$ von ihren Mittelwerten $\Rightarrow$ Positive Korrelation

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

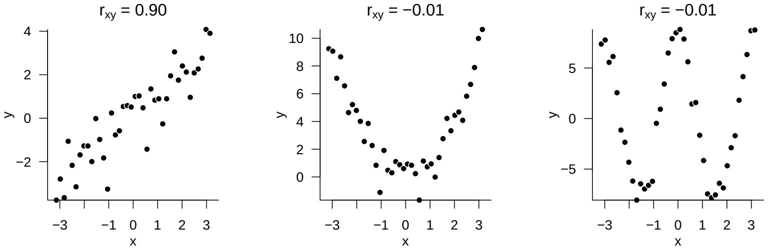
Keine häufigen richtungsgleichen oder -entgegengesetzten Abweichungen $\Rightarrow$ Keine Korrelation

# Korrelation

## Beispiele



Funktionale Abhängigkeiten und Stichprobenkorrelation




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

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

## Theorem (Kovarianz und Korrelation bei linear-affinen Transformationen)

$\xi$ und $\upsilon$ 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

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

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

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

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

$\xi$ und $\upsilon$ sind dann als Linearkombinationen von Zufallsvariablen selbst Zufallsvariablen. Die Korrelation von $\xi$ und $\upsilon$ 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 $\rho\left(a^T x, b^T y\right)$ und $\rho\left(\left(\alpha a^T\right) x, \left(\beta b^T\right) y\right)$ 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 $\rho(\xi, \upsilon) = \rho\left(a^T x, b^T y\right)$ 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.

## Grundzüge der Kanonischen Korrelationsanalyse

Da mit dem Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen die Varianzen zu verschiedenen skalaren Vielfachen von $a^T x$ und $b^T y$ 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), Kapitel 10. Dabei werden die Zufallsvektoren $x$ und $y$ in einem Zufallsvektor

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \tag{11}$$

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), 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 $A$ 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

$$A^{r/s} = Q\Lambda^{r/s}Q^T \text{ mit } \Lambda^{r/s} = \text{diag}\left(\lambda_i^{r/s}\right). \tag{12}$$

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

$$A^{1/2} = Q\Lambda^{1/2}Q^T \text{ mit } \Lambda^{1/2} = \text{diag}\left(\lambda_i^{1/2}\right). \tag{13}$$

Bemerkungen

- Offenbar gilt
$$\left(A^{1/2}\right)^2 = Q\Lambda^{1/2}Q^TQ\Lambda^{1/2}Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = Q\Lambda Q^T = A. \tag{14}$$
- Weiterhin gilt
$$\left(A^{-1/2}\right)^2 = Q\Lambda^{-1/2}Q^TQ\Lambda^{-1/2}Q^T = Q\Lambda^{-1/2}\Lambda^{-1/2}Q^T = Q\Lambda^{-1}Q^T = A^{-1}. \tag{15}$$
- Schließlich gilt
$$\begin{aligned} A^{-1/2}AA^{-1/2} = Q\Lambda^{-1/2}Q^TQ\Lambda Q^TQ\Lambda^{-1/2}Q^T &= Q\Lambda^{-1/2}\Lambda\Lambda^{-1/2}Q^T \\ &= Q\Lambda\Lambda^{-1}Q^T \\ &= I_m \end{aligned} \tag{16}$$

## 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 $\lambda$ von $AB$ $w := Bv$ ein Eigenvektor von $BA$ zum Eigenwert $\lambda$ 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 $\lambda_1$ sei der größte Eigenwert von $B^{-1}A$ mit assoziiertem Eigenvektor $v_1 \in \mathbb{R}^m$. Dann ist $\lambda_1$ eine Lösung des Optimierungsproblems

$$\max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \tag{17}$$

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

$$f : \mathbb{R}^m \to \mathbb{R}, x \mapsto f(x) := x^T A x, \tag{18}$$

dass

$$v_1 = \arg\max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1 \tag{19}$$

und dass

$$\lambda_1 = \max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \tag{20}$$

Beweis

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

$$y := B^{1/2}x \Leftrightarrow x = B^{-1/2}y \tag{21}$$

Dann kann mit der symmetrischen Matrix

$$K := B^{-1/2}AB^{-1/2} \in \mathbb{R}^{m \times m} \tag{22}$$

das Optimierungsproblem (17) geschrieben werden als

$$\max_y y^T K y \text{ unter der Nebenbedingung } y^T y = 1. \tag{23}$$

Dies gilt, weil

$$\max_x x^T Ax \Leftrightarrow \max_y \left(B^{-1/2}y\right)^T A \left(B^{-1/2}y\right) \Leftrightarrow \max_y y^T B^{-1/2}AB^{-1/2}y \Leftrightarrow \max_y y^T K y \tag{24}$$

und

$$x^T Bx = 1 \Leftrightarrow y^T B^{-1/2}BB^{-1/2}y = 1 \Leftrightarrow y^T y = 1. \tag{25}$$

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

$$K = Q\Lambda Q^T, \tag{26}$$

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

Beweis (fortgeführt)

Mit der orthogonalen Matrix $Q$ aus obiger Orthornomalzerlegung sei nun

$$z := Q^T y \Leftrightarrow y := Qz. \tag{27}$$

Dann kann das Optimierungsproblem (23) geschrieben werden als

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \text{ unter der Nebenbedingung } z^T z = 1, \tag{28}$$

weil

$$\max_y y^T K y \Leftrightarrow \max_z (Qz)^T K (Qz) \Leftrightarrow \max_z z^T Q^T Q \Lambda Q^T Q z \Leftrightarrow \max_z z^T \Lambda z \Leftrightarrow \max_z \sum_{i=1}^m \lambda_i z_i^2 \tag{29}$$

und

$$y^T y = 1 \Leftrightarrow (Qz)^T Qz = 1 \Leftrightarrow z^T Q^T Q z = 1 \Leftrightarrow z^T z = 1. \tag{30}$$

Beweis (fortgeführt)

Die Eigenwerte von $K$ seien nun absteigend sortiert, also $\lambda_1 \geq \cdots \geq \lambda_m$. Dann gilt für das Optimierungsproblem (28), dass

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_1, \tag{31}$$

weil

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \max_z \sum_{i=1}^m \lambda_1 z_i^2 = \lambda_1 \max_z \sum_{i=1}^m z_i^2 = \lambda_1 \tag{32}$$

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

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 = \lambda_1, \tag{33}$$

für $z := e_1 = (1, 0, ..., 0)^T$. Zusammenfassend heißt das, dass $z = e_1$ eine Lösung des Optimierungsproblem (28) ist und das $\lambda_1$ das entsprechende Maximum ist.

Beweis (fortgeführt)

Damit ergibt sich aber sofort, dass dann

$$y = Qz = Qe_1 = q_1 \text{ und } x = B^{-1/2}q_1 \tag{34}$$

Lösungen der äquivalenten Optimierungsprobleme (23) und (17), 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

$$B^{-1/2}B^{-1/2}A = B^{-1}A \tag{35}$$

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

$$\max_x x^T Ax \text{ unter der Nebenbedingung } x^T Bx = 1. \tag{36}$$

ist. $\square$

Korrelation

Algebraische Grundlagen

**Wahrscheinlichkeitstheoretische Grundlagen**

Modellformulierung

Modellschätzung

Selbstkontrollfragen

## Theorem (Kovarianzmatrizen von Zufallsvektoren)

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \tag{37}$$

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

$$\mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{38}$$

wobei

$$\begin{aligned} \Sigma_{xx} &:= \mathbb{E}\left(xx^T\right) \in \mathbb{R}^{m_x \times m_x} \\ \Sigma_{xy} &:= \mathbb{E}\left(xy^T\right) \in \mathbb{R}^{m_x \times m_y} \\ \Sigma_{yx} &:= \mathbb{E}\left(yx^T\right) \in \mathbb{R}^{m_y \times m_x} \\ \Sigma_{yy} &:= \mathbb{E}\left(yy^T\right) \in \mathbb{R}^{m_x \times m_y} \end{aligned} \tag{39}$$

Beweis

Nach Definition der Kovarianzmatrix eines Zufallsvektors gilt

$$
\begin{aligned}
\mathbb{C}(z) &= \mathbb{E}\left((z - \mathbb{E}(z))(z - \mathbb{E}(z))^T\right) \\
&= \mathbb{E}\left((z - 0_m)(z - 0_m)^T\right) \\
&= \mathbb{E}\left(zz^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} x^T & y^T \end{pmatrix}\right) \\
&= \mathbb{E}\left(\begin{pmatrix} xx^T & xy^T \\ yx^T & yy^T \end{pmatrix}\right) \\
&= \begin{pmatrix} \mathbb{E}\left(xx^T\right) & \mathbb{E}\left(xy^T\right) \\ \mathbb{E}\left(yx^T\right) & \mathbb{E}\left(yy^T\right) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}
\end{aligned}
\tag{40}
$$

$\square$

## Theorem (Linearkombinationen von Zufallsvektorpartitionen)

Es sei

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) = 0_m \text{ und } \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{41}$$

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

$$\xi := a^T x \text{ und } \upsilon := b^T y \tag{42}$$

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

Beweis von (1) und (2)

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

$$
\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left(\xi\xi\right) - \mathbb{E}(\xi)\mathbb{E}(\xi) \\
&= \mathbb{E}\left((a^Tx)(a^Tx)\right) - \mathbb{E}\left(a^Tx\right)\mathbb{E}\left(a^Tx\right) \\
&= \mathbb{E}\left((a^Tx)(a^Tx)^T\right) - \mathbb{E}\left(a^Tx\right)\mathbb{E}\left(a^Tx\right) \\
&= \mathbb{E}\left(a^Txx^Ta\right) - \mathbb{E}\left(a^Tx\right)\mathbb{E}\left(a^Tx\right) \\
&= a^T\mathbb{E}\left(xx^T\right)a - a^T\mathbb{E}(x)a^T\mathbb{E}(x) \\
&= a^T\mathbb{E}\left(xx^T\right)a - a^T0_{m_x}a^T0_{m_x} \\
&= a^T\Sigma_{xx}a.
\end{aligned}
\tag{43}
$$

Der Beweis zur Varianz von $\upsilon$ folgt dann analog.

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

$$
\begin{aligned}
\rho(\xi, \upsilon) &= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(\upsilon)}} \\
&= \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{1}\sqrt{1}} \\
&= \mathbb{C}(\xi, \upsilon) \\
&= \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) \\
&= \mathbb{E}\left((a^T x)(b^T y)\right) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\
&= \mathbb{E}\left((a^T x)(b^T y)^T\right) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\
&= \mathbb{E}\left(a^T x y^T b\right) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\
&= a^T \mathbb{E}\left(x y^T\right) b - a^T \mathbb{E}(x) b^T \mathbb{E}(y) \\
&= a^T \mathbb{E}\left(x y^T\right) b - a^T 0_{m_x} b^T 0_{m_y} \\
&= a^T \Sigma_{xy} b.
\end{aligned}
\tag{44}
$$

□

Korrelation

Algebraische Grundlagen

Wahrscheinlichkeitstheoretische Grundlagen

**Modellformulierung**

Modellschätzung

Selbstkontrollfragen

## Definition (Kanonische Koeffizientenvektoren, Variate, Korrelationen)

Es seien

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

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

$$K := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{46}$$

mit der Singulärwertzerlegung

$$K = A \Lambda B^T, \tag{47}$$

wobei

$$A := \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \end{pmatrix} \in \mathbb{R}^{m_x \times m_y} \text{ und } B := \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix} \in \mathbb{R}^{m_y \times m_y} \tag{48}$$

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

$$\Lambda := \operatorname{diag}\left(\lambda_1^{1/2}, ..., \lambda_k^{1/2}\right) \in \mathbb{R}^{m_y \times m_y}, \tag{49}$$

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

$$a_i := \Sigma_{xx}^{-1/2} \alpha_i \in \mathbb{R}^{m_x} \text{ und } b_i := \Sigma_{yy}^{-1/2} \beta_i \in \mathbb{R}^{m_y}. \tag{50}$$

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

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

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 $\xi, \upsilon_i$ und die kanonischen Korrelationen $\rho_i$ definiert wie oben. Dann gilt, dass für $1 \le r \le k$ das Maximum des $r$ten restringierten Optimierungsproblems

$$\phi_r = \max_{a,b} a^T \Sigma_{xy} b \quad (52)$$

unter den Nebenbedingungen

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

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

Bemerkungen

- $\phi_1$ 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 $\mathbb{V}(\upsilon) = b^T \Sigma_{yy} b = 1$
- $\phi_r$ 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 $\xi_i$

# Modellformulierung

Beweis

Wir betrachten das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a,b} \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ b^T \Sigma_{yy} b = 1,\ a_i^T \Sigma_{xx} a = 0, i = 1, \ldots, r-1 \tag{54}$$

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

$$\phi_r^2 = \max_a \left( \max_b \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N.} b^T \Sigma_{yy} b = 1 \right) \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ a_i^T \Sigma_{xy} a = 0, i = 1, \ldots, r-1 \tag{55}$$

von innen nach außen.

Schritt (1)

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

$$\max_b \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \tag{56}$$

Dieses Optimierungsproblem kann geschrieben werden als

$$\max_b b^T \Sigma_{yx} a a^T \Sigma_{xy} b \text{ u.d.N. } b^T \Sigma_{yy} b = 1, \tag{57}$$

weil gilt, dass

$$\left(a^T \Sigma_{xy} b\right)^2 = \left(a^T \Sigma_{xy} b\right) \left(a^T \Sigma_{xy} b\right) = \left(a^T \Sigma_{xy} b\right)^T a^T \Sigma_{xy} b = b^T \Sigma_{yx} a a^T \Sigma_{xy} b. \tag{58}$$

Beweis (fortgeführt)

Das Optimierungsproblem (57) kann nun mithilfe des Theorems zur Maximierung quadratischer Formen mit Nebenbedingungen gelöst werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yx} a a^T \Sigma_{xy} \text{ und } B := \Sigma_{yy}. \tag{59}$$

Dann hat (57) die Form

$$\max_b b^T A b \text{ unter der Nebenbedingung } b^T B b = 1, \tag{60}$$

Das Maximum von (60) entspricht nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen dem größten Eigenwert von

$$B^{-1} A = \Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy} \tag{61}$$

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

$$A := \Sigma_{yy}^{-1} \Sigma_{yx}, \quad b := a, \quad B := \Sigma_{xy} \tag{62}$$

und erhalten für den betreffenden Eigenwert

$$\lambda_a = b^T B A a = a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a. \tag{63}$$

als Lösung (Maximum) des restringierten Optimierungsproblems

$$\max_b \left(a^T \Sigma_{xy} b\right)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \tag{64}$$

Beweis (fortgeführt)

Schritt (2)

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

$$\phi_r^2 = \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1,\ a_i^T \Sigma_{xx} a = 0, i = 1, ..., r-1 \tag{65}$$

Dazu halten wir zunächst fest, dass (65) mit den Definitionen von $\alpha_i$ und $K$ in der Definition der Kanonischen Koeffizientenvektoren, Variaten, und Korrelationen geschrieben werden kann als

$$\phi_r^2 = \max_\alpha \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1,\ \alpha_i^T \alpha = 0, i = 1, ..., r-1, \tag{66}$$

denn

$$\begin{aligned}
\phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0 \Leftrightarrow \\
\phi_r^2 &= \max_\alpha a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 1, \alpha_i^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 0 \\
\phi_r^2 &= \max_\alpha \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\
\phi_r^2 &= \max_\alpha \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0 \\
\phi_r^2 &= \max_\alpha \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \alpha_i^T \alpha = 0
\end{aligned} \tag{67}$$

# Modellformulierung

Beweis (fortgeführt)

Dabei sind nach der betreffenden Definition die $\alpha_i$ 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 (66) der größte Eigenwert von $KK^T$ mit seinem assoziierten Eigenvektor. Die Nebenbedingung $\alpha_i^T \alpha = 0$ schränkt diese Wahl auf den $r$t-größten Eigenwert und seinen assoziierten Eigenvektor $\alpha_r$ ein. Mit der Definition von Eigenwerten und Eigenvektoren gilt also

$$\phi_r^2 = \alpha_r^T KK^T \alpha_r = \alpha_r^T \lambda_r \alpha_r = \lambda_r \alpha_r^T \alpha_r = \lambda_r. \tag{68}$$

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

$$K = A\Lambda B^T \Leftrightarrow KB = A\Lambda B^T B \Leftrightarrow KB = A\Lambda \Leftrightarrow K\beta_r = \alpha_r \lambda_r^{1/2} \tag{69}$$

dass

$$a_r^T \Sigma_{xy} b_r = \alpha_r^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \beta_r = \alpha_r^T K \beta_r = \alpha_r^T \alpha_r \lambda_r^{1/2} = \rho_r \tag{70}$$

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

$\square$

## Simulationsbeispiel

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

$$p(x) = N(x; 0_4, I_4) \text{ und } p(y|x) = N(y; Lx, G) \tag{71}$$

mit

$$L := \begin{pmatrix} 0.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 \end{pmatrix} \text{ und } G := \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \tag{72}$$

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

$$\begin{aligned} y_1 &= \phantom{-}x_3 + \varepsilon_1 \\ y_2 &= \phantom{-}x_1 + \varepsilon_2 \\ y_3 &= -x_4 + \varepsilon_3 \end{aligned} \tag{73}$$

mit

$$x_1 \sim N(0,1), x_3 \sim N(0,1), x_4 \sim N(0,1) \tag{74}$$

und

$$\varepsilon_1 \sim N(0,0.2), \varepsilon_2 \sim N(0,0.4), \varepsilon_3 \sim N(0,0.3) \tag{75}$$

# Modellformulierung

## Simulationsbeispiel

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

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N(0_7, \Sigma) \tag{76}$$

mit

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{77}$$

wobei

$$\Sigma_{xx} = I_4, \quad \Sigma_{xy} = L^T, \quad \Sigma_{yx} = L \text{ und } \Sigma_{yy} = G + LL^T. \tag{78}$$

Explizit ergibt sich also

$$\Sigma = \begin{pmatrix} I_4 & L^T \\ L & G + LL^T \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.2 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & 1.3 \end{pmatrix} \tag{79}$$

## Simulationsbeispiel

```r
# 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
```

## 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 $i = 1, \ldots, n$ seien

$$z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \text{ mit } \mathbb{E}(z_i) := 0_m \text{ und } \mathbb{C}(z_i) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{80}$$

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

$$C := \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{81}$$

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

$$\hat{a}_i := C_{xx}^{-1/2}\hat{\alpha}_i \in \mathbb{R}^{m_x}, \quad \hat{b}_i := C_{yy}^{-1/2}\hat{\beta}_i \in \mathbb{R}^{m_y} \text{ und } \hat{\rho}_i := \hat{\lambda}_i^{1/2} \tag{82}$$

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

$$\hat{K} := C_{xx}^{-1/2} C_{xy} C_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \tag{83}$$

$\hat{\alpha}_i$ 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.

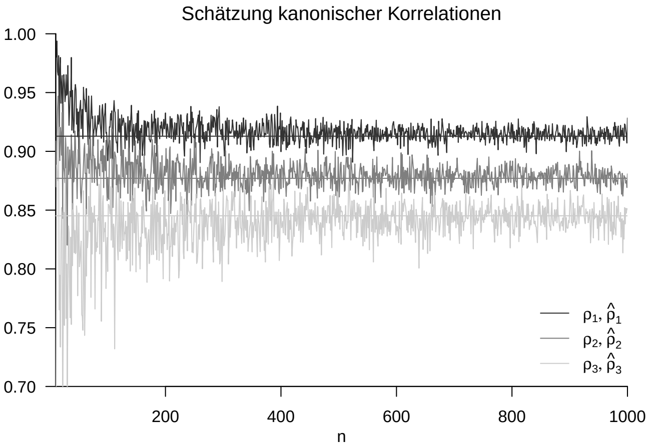
## Simulationsbeispiel

```
# R 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
```

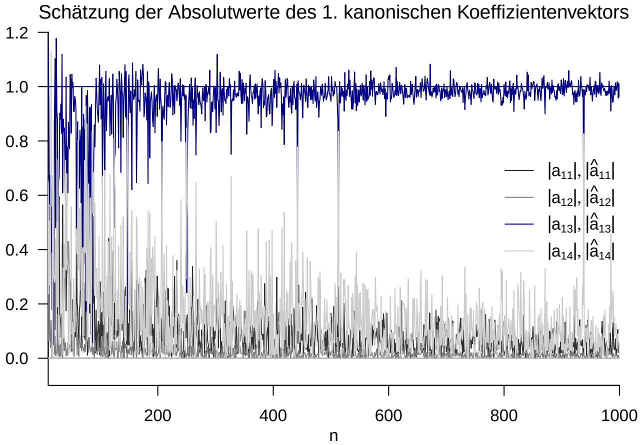
## Simulationsbeispiel

```
# 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]
}
```

## 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

## 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)
```

```
> rho_hat_1 :  0.995
> a_hat_1   :  -0.162 -0.174
> b_hat_1   :  -0.155 -0.0503
> rho_hat_2 :  0.501
> a_hat_2   :  -0.0603 0.312
> b_hat_2   :  -0.0813 0.777
```

## 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
```

## 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**

# Selbstkontrollfragen

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