Allgemeines Lineares Modell

BSc Psychologie, SoSe 2024

Joram Soch

(12) Partielle Korrelation

Semesterplan SoSe 2024: "Allgemeines Lineares Modell" & "Analyse und Dokumentation"

				Seminar	Vorlesung	Vorlesung "Allgemeines Lineares Modell" (B2)		Seminar "Analyse und Dokumentation" (C2)	
Wo.	KW	Jahr	ger. KW	Mittwoch	Donnerstag	Do, 13-15 (jede Wo.)	Do, 15-17 (ger. KW)	Mi, 11-13 (Gr. 1)	Mi, 13-15 (Gr. 2)
1	15	2024	nein	10.04.2024	11.04.2024	(1) Regression		(1) Ethik und ethische Formalitäten	
2	16	2024	ja	17.04.2024	18.04.2024	(2) Korrelation	(3) Matrizen	(2) Wissenschaftliche Berichte	
3	17	2024	nein	24.04.2024	25.04.2024	(3) Matrizen		(3) Offenheit und Transparenz	
4	18	2024	ja	01.05.2024	02.05.2024	(4) Normalverteilungen	(5) Modellformulierung	– Tag der Arbeit –	
5	19	2024	nein	08.05.2024	09.05.2024	– Christi Himmelfahrt –		(4) Einführung in Quarto	
6	20	2024	ja	15.05.2024	16.05.2024	(6) Parameterschätzung	(6) Parameterschätzung	(5) Offene Übung	
7	21	2024	nein	22.05.2024	23.05.2024	(7) T-Statistiken, (8) F-Statistiken		(6) Regression	
8	22	2024	ja	29.05.2024	30.05.2024	– Tag der Lehre –		(7) Korrelation	
9	23	2024	nein	05.06.2024	06.06.2024	(9) T-Tests		(8) Einstichproben-T-Test	
10	24	2024	ja	12.06.2024	13.06.2024	(10) Einfaktorielle Varianzanalyse	(11) Zweifaktorielle Varianzanalyse	(9) Zweistichproben-T-Test	
11	25	2024	nein	19.06.2024	20.06.2024	(11) Zweifaktorielle Varianzanalyse		(10) Einfaktorielle Varianzanalyse	
12	26	2024	ja	26.06.2024	27.06.2024	(12) Partielle Korrelation	(13) Multiple Regression	(11) Zweifaktorielle Varianzanalyse	
13	27	2024	nein	03.07.2024	04.07.2024	(13) Multiple Regression	(14) Kovarianzanalyse	(12) Multiple Regression	
14	28	2024	ja	10.07.2024	11.07.2024	Q & A		(13) Kovarianzanalyse	

Feiertag	Fleischmann	Soch

Motivation

Bedingte Korrelation

Partielle Korrelation

Selbstkontrollfragen

© 2024 Dirk Ostwald & Joram Soch, CC BY 4.0

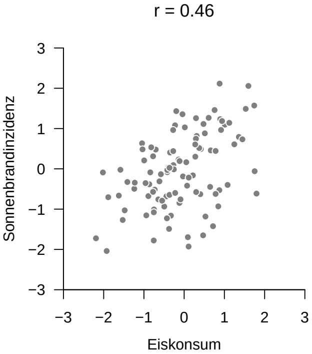
Motivation

Bedingte Korrelation

Partielle Korrelation

Selbstkontrollfragen

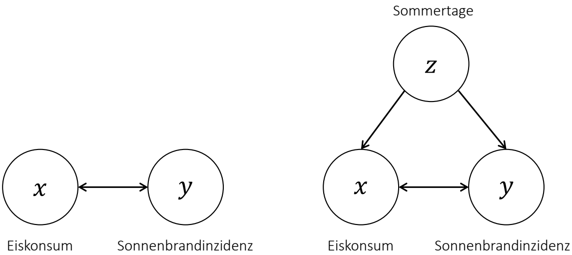
Jährlicher Eiskonsum und jährliche Sonnenbrandinzidenz



- Korrelation impliziert keine Kausalität.
- Kausalität wird zumeist als Koinzidenz mit zeitlicher Reihenfolge modelliert.
- Einstiege in die kausale Inferenz geben z.B. Pearl (2000) und Imbens and Rubin (2015).

Motivation

Jährlicher Eiskonsum und jährliche Sonnenbrandinzidenz



- Korrelation von Eiskonsum und Sonnenbrandinzidenz nach Korrektur für Sommertage?
- "Herausrechnen" des Einflusses von z auf die Kovariation von x und y?
- Im Folgenden bezeichnen wir <u>Zufallsvariablen</u> mit <u>kleinen lateinischen</u> Buchstaben.

$\Rightarrow$ bedingte Korrelation und partielle Korrelation im Falle dreier Zufallsvariablen

Motivation

Bedingte Korrelation

Partielle Korrelation

Selbstkontrollfragen

Bedingte Korelation

Definition (Bedingte Kovarianz und bedingte Korrelation)

Gegeben seien drei Zufallsvariablen x, y, z einer gemeinsamen Verteilung $\mathbb{P}(x, y, z)$.

Weiterhin sei $\mathbb{P}(x, y|z)$ die bedingte Verteilung von x und y, gegeben z. Dann heißt die Kovarianz von x und y in der Verteilung $\mathbb{P}(x, y|z)$ die *bedingte Kovarianz von x und y, gegeben z*, und wird mit $\mathbb{C}(x, y|z)$ bezeichnet.

Weiterhin seien $\mathbb{P}(x|z)$ und $\mathbb{P}(y|z)$ die marginalen Verteilungen von x bzw. y, gegeben z, und $\mathbb{S}(x|z)$ und $\mathbb{S}(y|z)$ seien die Standardabweichungen von x bzw. y hinsichtlich $\mathbb{P}(x|z)$ bzw. $\mathbb{P}(y|z)$.

Dann heißt die Korrelation von x und y in der Verteilung $\mathbb{P}(x, y|z)$

$$\rho(x, y|z) := \frac{\mathbb{C}(x, y|z)}{\mathbb{S}(x|z)\mathbb{S}(y|z)} \tag{1}$$

die *bedingte Korrelation von x und y, gegeben z*.

Bemerkungen

- Die bedingte Kovarianz zweier ZVen ist die Kovarianz dieser ZVen in einer bedingten Verteilung
- Die bedingte Korrelation zweier ZVen ist die Korrelation dieser ZVen in einer bedingten Verteilung.
- Durch Vertauschen der Variablennamen kann man analog $\rho(x, z|y)$ und $\rho(y, z|x)$ definieren.

Bedingte Korrelation

Beispiel

Die Zufallsvariablen x, y, z seien multivariat normalverteilt. Wir wollen die bedingte Korrelation von x und y gegeben z bestimmen. Für $v := (x, y, z)^{\mathrm{T}}$ gelte also, dass

$$v \sim N(\mu, \Sigma) \tag{2}$$

mit

$$\mu := \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 & \sigma_{x,z}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 & \sigma_{y,z}^2 \\ \sigma_{z,x}^2 & \sigma_{z,y}^2 & \sigma_z^2 \end{pmatrix} . \tag{3}$$

Um die Kovarianzmatrix der bedingten Verteilung von x und y, gegeben z zu bestimmen, definieren wir zunächst

$$\Sigma_{x,y} := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 \end{pmatrix}, \ \Sigma_z := \begin{pmatrix} \sigma_z^2 \end{pmatrix} \text{ und } \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^{\mathrm{T}} := \begin{pmatrix} \sigma_{x,z}^2 \\ \sigma_{y,z}^2 \end{pmatrix} , \tag{4}$$

sodass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix} \tag{5}$$

Mit dem Theorem zu bedingten Normalverteilungen (siehe Einheit (4) in *Allgemeines Lineares Modell*) ist dann die Kovarianzmatrix des Zufallsvektors $(x, y)^{\mathrm{T}}$ gegeben durch

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z}\Sigma_z^{-1}\Sigma_{z,(x,y)} . \tag{6}$$

Beispiel (fortgeführt)

Mit den Eigenschaften der multivariaten Normalverteilung gilt dann, dass die Diagonaleinträge von $\Sigma_{x,y|z}$ den bedingten Varianzen von x und y, gegeben z entsprechen und dass der Nichtdiagonaleintrag die bedingte Kovarianz von x und y, gegeben z ist. In anderen Worten gilt

$$\Sigma_{x,y|z} = \begin{pmatrix} \mathbb{C}(x,x|z) & \mathbb{C}(x,y|z) \\ \mathbb{C}(y,x|z) & \mathbb{C}(y,y|z) \end{pmatrix} . \tag{7}$$

Die bedingte Korrelation $\rho(x,y|z)$ von x und y, gegeben z ergibt sich dann aus den Einträgen von $\Sigma_{x,y|z}$ gemäß

$$\rho(x,y|z) = \frac{\mathbb{C}(x,y|z)}{\sqrt{\mathbb{C}(x,x|z)}\sqrt{\mathbb{C}(y,y|z)}} . \tag{8}$$

Beispiel (fortgeführt)

Es sei beispielsweise der Kovarianzmatrixparameter

$$\Sigma := \begin{pmatrix} 1.0 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.5 \\ 0.9 & 0.5 & 1.0 \end{pmatrix} \tag{9}$$

gegeben. Möchten wir nun die bedingte Korrelation $\rho(x, y|z)$ bestimmen, so evaluieren wir zunächst die Kovarianzmatrix des Zufallsvektors $(x, y)^{\mathrm{T}}$ gegeben z gemäß des in (6) aufgeschriebenen Theorems zu bedingten Normalverteilungen, indem wir entprechende in (4) defnierten Partitionierungen einsetzen.

$$\begin{aligned} \Sigma_{x,y|z} &= \Sigma_{x,y} - \Sigma_{(x,y),z}\Sigma_z^{-1}\Sigma_{z,(x,y)} \\ &= \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix} - \begin{pmatrix} 0.9 \\ 0.5 \end{pmatrix} \cdot 1 \cdot \begin{pmatrix} 0.9 & 0.5 \end{pmatrix} \\ &= \begin{pmatrix} 0.19 & 0.05 \\ 0.05 & 0.75 \end{pmatrix}. \end{aligned} \tag{10}$$

Bedingte Korrelation

Beispiel (fortgeführt)

Da, wie in Gleichung (7) beschrieben, die Einträge $\Sigma_{x,y|z}$ den entsprechenden bedingten Kovarianzen entsprechen, gilt

$$
\begin{aligned}
\Sigma_{x,y|z} &= \begin{pmatrix} 0.19 & 0.05 \\ 0.05 & 0.75 \end{pmatrix} \\
&= \begin{pmatrix} \mathbb{C}(x,x|z) & \mathbb{C}(x,y|z) \\ \mathbb{C}(y,x|z) & \mathbb{C}(y,y|z). \end{pmatrix}
\end{aligned}
\tag{11}
$$

Die bedingten Kovarianzen, die wir zur Bestimmung der bedingten Korrelation benötigen, sind entsprechend $\mathbb{C}(x,y|z) = 0.05$, $\mathbb{C}(x,x|z) = 0.19$ und $\mathbb{C}(y,y|z) = 0.75$. Somit ergibt sich die bedingte Korrelation als

$$
\begin{aligned}
\rho(x,y|z) &= \frac{\mathbb{C}(x,y|z)}{\sqrt{\mathbb{C}(x,x|z)}\sqrt{\mathbb{C}(y,y|z)}} \\
&= \frac{0.05}{\sqrt{0.19}\sqrt{0.75}} \approx 0.13.
\end{aligned}
\tag{12}
$$

Wir erhalten also

$$
\rho(x,y) = 0.50 \text{ und } \rho(x,y|z) \approx 0.13\ . \tag{13}
$$

Bedingte Korrelation

Beispiel (fortgeführt)

```
# Bedingte Korrelation bei Normalverteilung
S         = matrix(c(  1, 0.5, 0.9,                               # \Sigma
                     0.5,   1, 0.5,
                     0.9, 0.5,   1), nrow = 3, byrow = TRUE)
rho_xy    = S[1,2]/(sqrt(S[1,1])*sqrt(S[2,2]))                    # \rho(x,y)
S_xy_z    = S[1:2,1:2] - S[1:2,3] %*% solve(S[3,3]) %*% S[3,1:2] # \Sigma_{x,y|z}
Cov_xy_z  = S_xy_z[1,2]                                           # \mathbb{C}(x,y|z)
Cov_x_z   = S_xy_z[1,1]                                           # \mathbb{C}(x,x|z)
Cov_y_z   = S_xy_z[2,2]                                           # \mathbb{C}(y,y|z)
rho_xy_z  = Cov_xy_z/(sqrt(Cov_x_z)*sqrt(Cov_y_z))                # \rho(x,y|z)

# Ausgabe der (bedingten) Korrelationen
cat(  "rho(x,y)   :", rho_xy,
    "\nrho(x,y|z) :", rho_xy_z)
```

```
> rho(x,y)   : 0.5
> rho(x,y|z) : 0.132
```

Bedingte Korrelation

Theorem (Bedingte Korrelation und Korrelationen bei Normalverteilung)

x, y, z seien drei gemeinsam multivariat normalverteilte Zufallsvariablen. Dann gilt:

$$\rho(x, y|z) = \frac{\rho(x, y) - \rho(x, z)\rho(y, z)}{\sqrt{1 - \rho(x, z)^2}\sqrt{1 - \rho(y, z)^2}} \ . \tag{14}$$

Bemerkungen

- $\rho(x, y|z)$ kann bei Normalverteilung aus den Korrelationen $\rho(x, y), \rho(x, z), \rho(y, z)$ berechnet werden.
- Ein entsprechender Schätzer für $\rho(x, y|z)$ ergibt sich mit den Stichprobenkorrelationen $r_{x,y}, r_{x,z}, r_{y,z}$ als

$$r_{x,y|z} = \frac{r_{x,y} - r_{x,z} r_{y,z}}{\sqrt{1 - r_{x,z}^2}\sqrt{1 - r_{y,z}^2}} \tag{15}$$

Bedingte Korrelation

<u>Beweis</u>

Ohne Beschränkung der Allgemeinheit betrachten wir den Fall eines standardisierten multivariat normalverteilten Zufallsvektors $v := (x, y, z)^{\mathrm{T}}$ mit Kovarianzmatrixparameter

$$\Sigma := \begin{pmatrix} 1 & \rho(x,y) & \rho(x,z) \\ \rho(y,x) & 1 & \rho(y,z) \\ \rho(z,x) & \rho(z,y) & 1 \end{pmatrix}. \tag{16}$$

Wir definieren nun zunächst

$$\Sigma_{x,y} := \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix}, \Sigma_z := \begin{pmatrix} 1 \end{pmatrix} \text{ und } \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^{\mathrm{T}} := \begin{pmatrix} \rho(x,z) \\ \rho(y,z) \end{pmatrix}, \tag{17}$$

sodass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix}. \tag{18}$$

Mit dem Theorem zu bedingten Normalverteilungen (siehe Einheit (4) in *Allgemeines Lineares Modell*) ist dann die Kovarianzmatrix des Zufallsvektors (x, y) gegeben durch

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z}\Sigma_z^{-1}\Sigma_{z,(x,y)}. \tag{19}$$

Beweis (fortgeführt)

Es ergibt sich also

$$
\begin{aligned}
\begin{pmatrix} \sigma^2_{x,x|z} & \sigma^2_{x,y|z} \\ \sigma^2_{y,x|z} & \sigma^2_{y,y|z} \end{pmatrix} & = \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x,z) \\ \rho(y,z) \end{pmatrix} \Big(1\Big)^{-1} \Big(\rho(x,z) \quad \rho(y,z)\Big) \\
& = \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x,z)\rho(x,z) & \rho(x,z)\rho(y,z) \\ \rho(y,z)\rho(x,z) & \rho(y,z)\rho(y,z) \end{pmatrix} \\
& = \begin{pmatrix} 1 - \rho(x,z)^2 & \rho(x,y) - \rho(x,z)\rho(y,z) \\ \rho(y,x) - \rho(y,z)\rho(x,z) & 1 - \rho(y,z)^2 \end{pmatrix}.
\end{aligned}
\tag{20}
$$

Damit folgt dann direkt

$$
\rho(x,y|z) = \frac{\sigma^2_{x,y|z}}{\sqrt{\sigma^2_{x,x|z}}\sqrt{\sigma^2_{y,y|z}}} = \frac{\rho(x,y) - \rho(x,z)\rho(y,z)}{\sqrt{1-\rho(x,z)^2}\sqrt{1-\rho(y,z)^2}}. \tag{21}
$$

$\square$

Motivation

Bedingte Korrelation

Partielle Korrelation

Selbstkontrollfragen

Definition (Partielle Korrelation)

x, y, z seien Zufallsvariablen mit linear-affinen Abhängigkeiten zwischen x und z sowie zwischen y und z

$$
\begin{aligned}
x &:= \beta_0^{(x)} + \beta_1^{(x)} z + e^{(x)} \\
y &:= \beta_0^{(y)} + \beta_1^{(y)} z + e^{(y)} \ ,
\end{aligned}
\tag{22}
$$

sodass die Residualvariablen gegeben sind als

$$
\begin{aligned}
e^{(x)} &= x - \beta_0^{(x)} - \beta_1^{(x)} z \\
e^{(y)} &= y - \beta_0^{(y)} - \beta_1^{(y)} z \ .
\end{aligned}
\tag{23}
$$

Dann ist die *partielle Korrelation von x und y mit auspartialisiertem z* definiert als

$$
\rho(x, y \setminus z) := \rho\left(e^{(x)}, e^{(y)}\right). \tag{24}
$$

Bemerkungen

- $e^{(x)}$ ist die Zufallsvariable x, aus der der Einfluss von z "herausgerechnet" wurde.
- $e^{(y)}$ ist die Zufallsvariable y, aus der der Einfluss von z "herausgerechnet" wurde.
- $\rho(x, y \setminus z)$ ist also die Korrelation von x und y, wobei jeweils der Einfluss von z "herausgerechnet" wurde.

Partielle Korrelation

Definition (Partielle Stichprobenkorrelation)

x, y, z seien Zufallsvariablen mit linear-affinen Abhängigkeiten zwischen x und z sowie zwischen y und z wie in der Definition der partiellen Korrelation. Weiterhin seien

- $\{(x_i, y_i, z_i)\}_{i=1,...,n}$ eine Menge von Realisierungen des Zufallsvektors $(x, y, z)^{\mathrm{T}}$,
- $\hat{\beta}_0^{(x)}, \hat{\beta}_1^{(x)}$ die Ausgleichsgeradenparameter für $\{(x_i, z_i)\}_{i=1,...,n}$,
- $\hat{\beta}_0^{(y)}, \hat{\beta}_1^{(y)}$ die Ausgleichsgeradenparameter für $\{(y_i, z_i)\}_{i=1,...,n}$.

Schließlich seien für $i = 1, ..., n$

- $e_i^{(x)} := x_i - \hat{\beta}_0^{(x)} - \hat{\beta}_1^{(x)} z_i$
- $e_i^{(y)} := y_i - \hat{\beta}_0^{(y)} - \hat{\beta}_1^{(y)} z_i$

die Residuen der jeweiligen Ausgleichsgeraden. Dann heißt die Stichprobenkorrelation der Wertemenge $\left\{\left(e_i^{(y)}, e_i^{(x)}\right)\right\}_{i=1,...,n}$ *partielle Stichprobenkorrelation der x_i und y_i mit auspartialisierten z_i*.

Bemerkungen

- Die partielle Stichprobenkorrelation wird als Schätzer der partiellen Korrelation genutzt.

Theorem (Bedingte und partielle Korrelation bei Normalverteilung)

x, y, z seien drei gemeinsam multivariat normalverteilte Zufallsvariablen. Dann gilt:

$$\rho(x, y|z) = \rho(x, y \setminus z) \; . \tag{25}$$

Bemerkungen

- Generell sind bedingte und partielle Korrelationen nicht identisch.
- Für Details, siehe zum Beispiel Lawrance (1976) und Baba, Shibata, and Sibuya (2004).
- Mit der Gleichheit von bedingter und partieller Korrelation bei Normalverteilung ist der Schätzer für die bedingte Korrelation auch ein Schätzer für die partielle Korrelation bei gemeinsam normalverteilten Zufallsvariablen.

© 2024 Dirk Ostwald & Joram Soch, CC BY 4.0

<u>Beweis</u>

Ohne Beschränkung der Allgemeinheit betrachten wir erneut den Fall eines standardisierten multivariat normalverteilten Zufallsvektors $v := (x, y, z)^{\mathrm{T}}$ mit Kovarianzmatrixparameter

$$\Sigma := \begin{pmatrix} 1 & \rho(x,y) & \rho(x,z) \\ \rho(y,x) & 1 & \rho(y,z) \\ \rho(z,x) & \rho(z,y) & 1 \end{pmatrix} . \tag{26}$$

Dann definieren wir

$$A := \begin{pmatrix} 1 & 0 & -\beta_1^{(x)} \\ 0 & 1 & -\beta_1^{(y)} \end{pmatrix} \quad \text{und} \quad b := \begin{pmatrix} -\beta_0^{(x)} \\ -\beta_0^{(y)} \end{pmatrix} , \tag{27}$$

sodass gilt

$$\begin{pmatrix} e^{(x)} \\ e^{(y)} \end{pmatrix} = Av + b = \begin{pmatrix} 1 & 0 & -\beta_1^{(x)} \\ 0 & 1 & -\beta_1^{(y)} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} -\beta_0^{(x)} \\ -\beta_0^{(y)} \end{pmatrix} = \begin{pmatrix} x - \beta_0^{(x)} - \beta_1^{(x)} z \\ y - \beta_0^{(y)} - \beta_1^{(y)} z \end{pmatrix} . \tag{28}$$

Die partielle Korrelation ergibt sich dann als

$$\rho(x, y \setminus z) = \rho(e^{(x)}, e^{(y)}) . \tag{29}$$

© 2024 Dirk Ostwald & Joram Soch, CC BY 4.0

Partielle Korrelation

Beweis (fortgeführt)

Mit dem Theorem zur linear-affinen Transformation von multivariaten Normalverteilungen (siehe Einheit (4) in *Allgemeines Lineares Modell*) gilt für den Vektor der Residualvariablen

$$v \sim N(\mu, \Sigma) \quad \Rightarrow \quad \begin{pmatrix} e^{(x)} \\ e^{(y)} \end{pmatrix} = Av + b \sim N\left(A\mu + b, A\Sigma A^{\mathrm{T}}\right) . \tag{30}$$

Der Kovarianzmatrixparameter der resultierenden multivariaten Normalverteilung ergibt sich also zu

$$\begin{aligned}
A\Sigma A^{\mathrm{T}} &= \begin{pmatrix} 1 & 0 & -\beta_1^{(x)} \\ 0 & 1 & -\beta_1^{(y)} \end{pmatrix} \begin{pmatrix} 1 & \rho(x,y) & \rho(x,z) \\ \rho(y,x) & 1 & \rho(y,z) \\ \rho(z,x) & \rho(z,y) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\beta_1^{(x)} & -\beta_1^{(y)} \end{pmatrix} \\
&= \begin{pmatrix} 1 - \beta_1^{(x)}\rho(z,x) & \rho(x,y) - \beta_1^{(x)}\rho(z,y) & \rho(x,z) - \beta_1^{(x)} \\ \rho(y,x) - \beta_1^{(y)}\rho(z,x) & 1 - \beta_1^{(y)}\rho(z,y) & \rho(y,z) - \beta_1^{(y)} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\beta_1^{(x)} & -\beta_1^{(y)} \end{pmatrix} \\
&= \begin{pmatrix} 1 - \beta_1^{(x)}\rho(z,x) - \beta_1^{(x)}\rho(x,z) + \left(\beta_1^{(x)}\right)^2 & \rho(x,y) - \beta_1^{(x)}\rho(z,y) - \beta_1^{(y)}\rho(x,z) + \beta_1^{(x)}\beta_1^{(y)} \\ \rho(y,x) - \beta_1^{(y)}\rho(z,x) - \beta_1^{(x)}\rho(y,z) + \beta_1^{(x)}\beta_1^{(y)} & 1 - \beta_1^{(y)}\rho(z,y) - \beta_1^{(y)}\rho(y,z) + \left(\beta_1^{(y)}\right)^2 \end{pmatrix} \\
&= \begin{pmatrix} 1 - 2\beta_1^{(x)}\rho(x,z) + \left(\beta_1^{(x)}\right)^2 & \rho(x,y) - \beta_1^{(x)}\rho(z,y) - \beta_1^{(y)}\rho(x,z) + \beta_1^{(x)}\beta_1^{(y)} \\ \rho(y,x) - \beta_1^{(y)}\rho(z,x) - \beta_1^{(x)}\rho(y,z) + \beta_1^{(x)}\beta_1^{(y)} & 1 - 2\beta_1^{(y)}\rho(y,z) + \left(\beta_1^{(y)}\right)^2 \end{pmatrix} .
\end{aligned} \tag{31}$$

Partielle Korrelation

Beweis (fortgeführt)

Besteht ein linear-affiner Zusammenhang der Form $x := \beta_0 + \beta_1 y + e$ zwischen zwei Zufallsvariablen, dann gilt für die Korrelation dieser Zufallsvariablen:

$$\rho(x,y) = \frac{\mathbb{S}(x)}{\mathbb{S}(y)}\beta_1 \quad \Leftrightarrow \quad \beta_1 = \frac{\mathbb{S}(y)}{\mathbb{S}(x)}\rho(x,y)\,. \tag{32}$$

Wir halten dies hier ohne Beweis fest, bemerken aber, dass analog auch in der Stichprobe für den Steigungsparameter der Ausgleichsgerade gilt:

$$\hat{\beta}_1 = \frac{c_{xy}}{s_x^2}, \quad r_{xy} = \frac{c_{xy}}{s_x s_y} \quad \Rightarrow \quad \hat{\beta}_1 = \frac{s_y}{s_x} r_{xy}\,. \tag{33}$$

Im vorliegenden Fall der standardisierten Zufallsvariablen x, y und z mit Varianz und Standardabweichung gleich 1 ergibt sich

$$\beta_1^{(x)} = \frac{1}{1}\rho(x,z) = \rho(x,z) \quad \text{und} \quad \beta_1^{(y)} = \frac{1}{1}\rho(y,z) = \rho(y,z) \tag{34}$$

und damit

$$A\Sigma A^{\mathrm{T}} = \begin{pmatrix} 1-\rho(x,z)^2 & \rho(x,y)-\rho(x,z)\rho(y,z) \\ \rho(x,y)-\rho(x,z)\rho(y,z) & 1-\rho(y,z)^2 \end{pmatrix} = \begin{pmatrix} \sigma^2_{x,x\backslash z} & \sigma^2_{x,y\backslash z} \\ \sigma^2_{y,x\backslash z} & \sigma^2_{y,y\backslash z} \end{pmatrix}\,. \tag{35}$$

Partielle Korrelation

Beweis (fortgeführt)

Schließlich folgt die partielle Korrelation mit

$$
\begin{aligned}
\rho(x, y \setminus z) &= \rho(e^{(x)}, e^{(y)}) \\
&= \frac{\sigma^2_{x,y\setminus z}}{\sqrt{\sigma^2_{x,x\setminus z}}\sqrt{\sigma^2_{y,y\setminus z}}} \\
&= \frac{\rho(x,y) - \rho(x,z)\rho(y,z)}{\sqrt{1-\rho(x,z)^2}\sqrt{1-\rho(y,z)^2}} \; .
\end{aligned} \tag{36}
$$

Dies entspricht der Formel im Theorem zur bedingten Korrelation bei Normalverteilung. Mithin sind die bedingte Korrelation von x und y, gegeben z, und die partielle Korrelation von x und y mit herauspartialisiertem z, identisch, wenn x, y und z einer gemeinsamen Normalverteilung folgen. $\square$

Partielle Korrelation

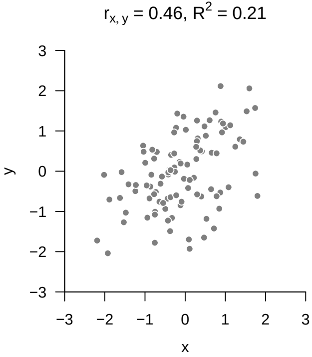
Beispiel

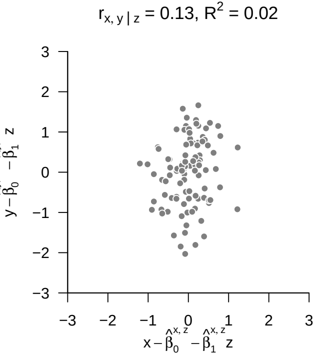
```
# Modellformulierung und Datenrealisierung
library(MASS)                                         # multivariate Normalverteilung
set.seed(1)                                           # reproduzierbare Daten
S     = matrix(c(  1,0.5,0.9,                        # Kovarianzmatrixparameter \Sigma
                 0.5,  1,0.5,
                 0.9,0.5,  1), nrow = 3, byrow = TRUE)
n     = 1e6                                           # Anzahl Realisierungen von v := (x,y,z)^T
xyz   = mvrnorm(n,rep(0,3),S)                         # Realisierungen        von v := (x,y,z)^T

# partielle Stichprobenkorrelation als Stichprobenkorrelation der Residuen
bars  = apply(xyz, 2, mean)                           # Stichprobenmittel
s     = apply(xyz, 2, sd)                             # Stichprobenstandardabweichungen
c     = cov(xyz)                                      # Stichprobenkovarianzen
b_xz1 = c[1,3]/c[3,3]                                 # beta_1 (x,z)
b_xz0 = bars[1] - b_xz1*bars[3]                       # beta_0 (x,z)
b_yz1 = c[2,3]/c[3,3]                                 # beta_1 (y,z)
b_yz0 = bars[2] - b_yz1*bars[3]                       # beta_0 (y,z)
e_xz  = xyz[,1] - b_xz1*xyz[,3] - b_xz0               # Residuen e^{(x)}
e_yz  = xyz[,2] - b_yz1*xyz[,3] - b_yz0               # Residuen e^{(y)}
pr_e  = cor(e_xz,e_yz)                                # \rho(x,y\z)

# partielle Stichprobenkorrelation als bedingte Stichprobenkorrelation
r      = cor(xyz)                                     # Stichprobenkorrelationsmatrix
pr_r_n = r[1,2]-r[1,3]*r[2,3]                         # \rho(x,y\z) Formel Zähler
pr_r_d = sqrt((1-r[1,3]^2)*(1-r[2,3]^2))              # \rho(x,y\z) Formel Nenner
pr_r   = pr_r_n/pr_r_d                                # \rho(x,y\z)

# partielle Stichprobenkorrelation mithilfe der Funktion pcor() aus der Toolbox "ppcor"
library(ppcor)                                        # Laden der Toolbox
pr_t   = pcor(xyz)                                    # \rho(x,y\z),\rho(x,z\y),\rho(y,z\x)

# Ausgabe
cat(  "r(x,y)                            :",  r[1,2],
    "\nr(x,y\\z) aus Residuenkorrelation :", pr_e,
    "\nr(x,y\\z) aus Korrelationen       :", pr_r,
    "\nr(x,y\\z) aus Toolbox             :", pr_t$estimate[1,2])
```

```
> r(x,y)                           : 0.5
> r(x,y\z) aus Residuenkorrelation : 0.133
> r(x,y\z) aus Korrelationen       : 0.133
> r(x,y\z) aus Toolbox             : 0.133
```

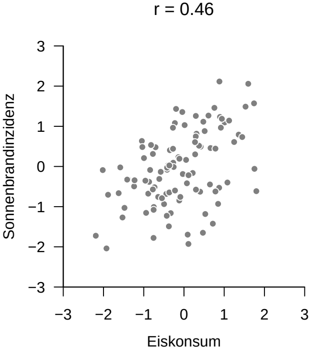
© 2024 Dirk Ostwald & Joram Soch, CC BY 4.0

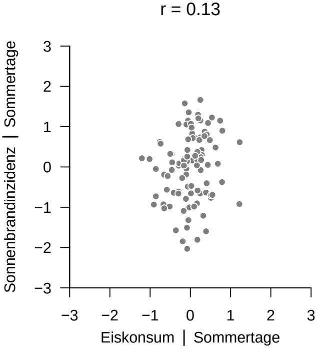
Partielle Korrelation





Partielle Korrelation





Motivation

Bedingte Korrelation

Partielle Korrelation

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die Motivation zur Bestimmung bedingter und partieller Korrelationen.
2. Definieren Sie die Begriffe der bedingten Kovarianz und der bedingten Korrelation.
3. Geben Sie das Theorem zu bedingter Korrelation und Korrelationen bei Normalverteilung wieder.
4. Definieren Sie den Begriff der partiellen Korrelation.
5. Definieren Sie den Begriff der partiellen Stichprobenkorrelation.
6. Geben Sie das Theorem zu bedingter und partieller Korrelation bei Normalverteilung wieder.
7. Erläutern Sie die Berechnung einer partiellen Korrelation anhand eines Anwendungsbeispiels.

Baba, Kunihiro, Ritei Shibata, and Masaaki Sibuya. 2004. "Partial Correlation and Conditional Correlation as Measures of Conditional Independence." *Australian & New Zealand Journal of Statistics* 46 (4): 657–64. https://doi.org/10.1111/j.1467-842X.2004.00360.x.

Imbens, Guido, and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Academic Press.

Lawrance, A. J. 1976. "On Conditional and Partial Correlation." *The American Statistician* 30 (3): 146. https://doi.org/10.2307/2683864.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K. ; New York: Cambridge University Press.