



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(0) Formalia

Modul B2 Inferenzstatistik | Allgemeines Lineares Modell

- Freitags 9 - 12 Uhr (3 SWS), G22A-013 (70 Pl.)
- Kursmaterialien (Folien, Videos) auf der [Kurswebseite](#)
- Ankündigungen über die [Moodleseite](#)
- Benotete Multiple Choice Klausur (30 Fragen) Ende Sommersemester 2022
- Klausurwiederholungstermin am Ende des Wintersemesters 2022/23
- Klausurtermin und Klausurort gemäß Prüfungsplan des [FNW Prüfungsamtes](#)

Modul B2 Inferenzstatistik | Allgemeines Lineares Modell

Datum	Einheit	Thema
08.04.2022	Grundlagen Osterpause	(1) Regression
22.04.2022	Grundlagen	(2) Korrelation
29.04.2022	Grundlagen	(3) Matrizen
06.05.2022	Grundlagen	(4) Normalverteilungen
13.05.2022	Theorie	(5) Modellformulierung
20.05.2022	Theorie	(6) Modellschätzung
27.05.2022	Theorie	(7) Modellevaluation
03.06.2021	Anwendung	(8) Studiendesign
10.06.2021	Anwendung	(9) T-Tests
17.06.2021	Anwendung	(10) Einfaktorielle Varianzanalyse
24.06.2022	Anwendung	(11) Zweifaktorielle Varianzanalyse
01.07.2022	Anwendung	(12) Multiple Regression
08.07.2022	Anwendung	(13) Kovarianzanalyse
Juli 2022	Klausurtermin	
März 2023	Klausurwiederholungstermin	

Modul B2 Inferenzstatistik | Allgemeines Lineares Modell

Themenkomplexe laut Modulhandbuch BSc Psychologie 11/2019

Deskription, grafische und tabellarische Darstellung von Daten	PDS
Verteilungskennwerte: Maße der zentralen Tendenz, Streuungsmaße	PDS
Wahrscheinlichkeitstheorie und Wahrscheinlichkeitsrechnung	WTFI
Stichprobe und Grundgesamtheit	WTFI
Formulierung und Prüfung von Hypothesen	WTFI
Effektstärke und optimale Stichprobenumfänge	WTFI
Verfahren zur Prüfung von Unterschiedshypothesen	ALM
Verfahren zur Prüfung von Zusammenhangshypothesen	ALM
Korrelation und multiple Regression	ALM
Varianzanalyse	ALM
Anwendungsbeispiele	PDS, WTFI, ALM

Einführung in die Faktorenanalyse ⇒ **Multivariate Datenanalyse**

Modul B2 Inferenzstatistik | Allgemeines Lineares Modell

- Vorlesungsfolien inklusive Selbstkontrollfragen sind klausurrelevant
- Altklausuren finden sich auf den Kurswebseiten früherer Jahre
- Als weiterführende Literatur bieten sich an
 - Werner, J. (1997) Lineare Statistik
 - Searle, S.R. (1971) Linear Models
 - Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., Tutz, G. (2016) Statistik
 - DeGroot, M.H. & Shervish, M.J. (2012) Probability and Statistics
 - Fox, J., Sanford, W. (2019) An R Companion to Applied Regression

Lernphasen

Phase 1: Überblicken

- Überblick durch Vorlesung und Überfliegen der Materialien.
- Verstehen einfacher Zusammenhänge.
- Verstehen, was man nicht versteht.

Phase 2: Verstehen

- Erarbeiten des Verstehens komplexer Zusammenhänge.
- Schriftliche Beantwortung der Selbstkontrollfragen.
- Klärung von Details, z.B. mit Vorlesungsvideos.

Phase 3: Memorisieren

- Auswendiglernen aller Inhalte.
- Aktive Wiedergabe der Inhalte, schriftlich oder mündlich.
- Teilnahme an der Klausur.

Teilen Sie große Aufgaben immer in viele kleine, gut zu bewältigende Aufgaben!

Sie machen Schreibtischarbeit, treiben Sie also täglich Sport!

Q & A



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(1) Regression

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Anwendungsszenario

Psychotherapie



Mehr Therapiestunden

⇒ Höhere Wirksamkeit?

Unabhängige Variable

- Anzahl Therapiestunden

Abhängige Variable

- Symptomreduktion

Methode der kleinsten Quadrate

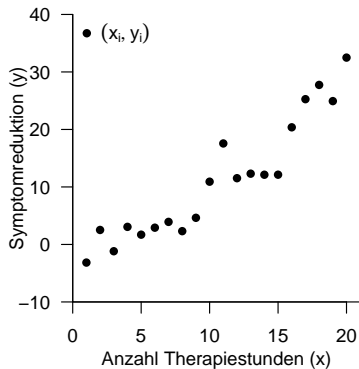
Beispieldatensatz

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

y_i	x_i
-3.15	1
2.52	2
-1.18	3
3.06	4
1.70	5
2.91	6
3.92	7
2.31	8
4.63	9
10.91	10
17.56	11
11.52	12
12.31	13
12.12	14
12.13	15
20.37	16
25.26	17
27.75	18
24.93	19
32.49	20

Methode der kleinsten Quadrate

Beispieldatensatz



Welcher funktionaler Zusammenhang zwischen x und y liegt den Daten zugrunde?

Definition (Ausgleichsgerade)

Für $\beta := (\beta_0, \beta_1)^T \in \mathbb{R}^2$ heißt die linear-affine Funktion

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \beta_0 + \beta_1 x, \quad (1)$$

für die für eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ die Funktion

$$q : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^n (y_i - f_\beta(x_i))^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (2)$$

der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$ ihr Minimum annimmt, die *Ausgleichsgerade für die Wertemenge* $\{(x_1, y_1), \dots, (x_n, y_n)\}$.

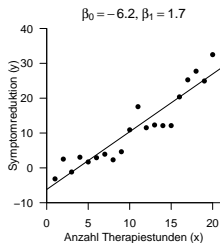
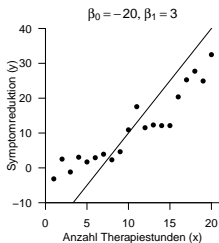
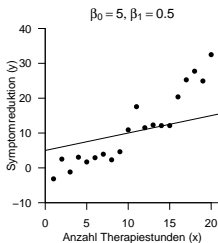
Bemerkungen

- Wir nehmen hier ohne Beweis an, dass das Minimum von q eindeutig ist.

Methode der kleinsten Quadrate

Linear-affine Funktionen $f_{\beta}(x) := \beta_0 + \beta_1 x$

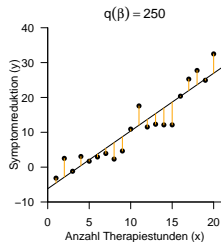
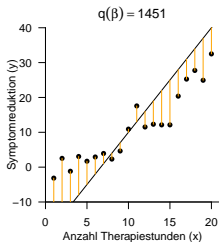
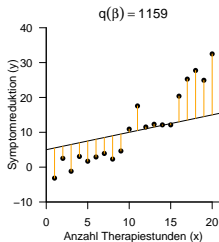
- β_0 : Schnittpunkt von Gerade und y -Achse (“Offset Parameter”)
- β_1 : y -Differenz pro x -Einheitsdifferenz (“Steigungsparameter”)



Methode der kleinsten Quadrate

Funktion der quadrierten vertikalen Abweichungen

$$q(\beta) := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (3)$$



— $y_i - (\beta_0 + \beta_1 x_i)$ für $i = 1, \dots, n$

Theorem (Ausgleichsgerade)

Für eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ hat die Ausgleichsgerade die Form

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \hat{\beta}_0 + \hat{\beta}_1 x, \quad (4)$$

wobei mit der Stichprobenkovarianz c_{xy} der (x_i, y_i) -Werte, der Stichprobenvarianz s_x^2 der x_i -Werte und den Stichprobenmitteln \bar{x} und \bar{y} der x_i - und y_i -Werte, respektive, gilt, dass

$$\hat{\beta}_1 = \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (5)$$

Bemerkungen

- Mit den Definitionen von c_{xy} und s_x^2 gilt also

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

- Man spricht hier von der Stichprobenkovarianz c_{xy} , auch wenn die Werte x_1, \dots, x_n oft nicht als Realisierungen einer Stichprobe X_1, \dots, X_n verstanden werden, sondern als gegebene oder selbst gewählte Zahlen.

Methode der kleinsten Quadrate

Beweis

Wir betrachten die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f(x_i)$ als Funktion von β_0 und β_1 und bestimmen Werte $\hat{\beta}_0$ und $\hat{\beta}_1$, für die diese Funktion ihr Minimum annimmt, die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f(x_i)$ also minimal ist. Wir betrachten also die Funktion

$$q : \mathbb{R}^2 \rightarrow \mathbb{R}, (\beta_0, \beta_1) \mapsto q(\beta_0, \beta_1) := \sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2. \quad (7)$$

Um das Minimum dieser Funktion zu bestimmen, berechnen wir zunächst die partiellen Ableitungen hinsichtlich β_0 und β_1 und setzen diese gleich 0. Es ergibt sich zunächst

$$\begin{aligned} \frac{\partial}{\partial \beta_0} q(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_0} \left(\sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2 \right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_0} \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2 \\ &= \sum_{i=1}^n 2 \left(y_i - (\beta_0 + \beta_1 x_i) \right) \frac{\partial}{\partial \beta_0} \left(y_i - \beta_0 - \beta_1 x_i \right) \\ &= -2 \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i \right) \end{aligned} \quad (8)$$

Methode der kleinsten Quadrate

Beweis (fortgeführt)

Weiterhin ergibt sich

$$\begin{aligned}\frac{\partial}{\partial \beta_1} q(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_1} \left(\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} (y_i - (\beta_0 + \beta_1 x_i))^2 \\ &= \sum_{i=1}^n 2 (y_i - (\beta_0 + \beta_1 x_i)) \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 x_i) \\ &= -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i\end{aligned}\tag{9}$$

Nullsetzen beider partieller Ableitungen ergibt dann

$$\begin{aligned}\frac{\partial}{\partial \beta_0} q(\beta_0, \beta_1) = 0 \text{ und } \frac{\partial}{\partial \beta_1} q(\beta_0, \beta_1) = 0 \\ \Leftrightarrow -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \text{ und } -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \\ \Leftrightarrow \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \text{ und } \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0\end{aligned}\tag{10}$$

Methode der kleinsten Quadrate

Beweis (fortgeführt)

und weiter

$$\sum_{i=1}^n y_i - \sum_{i=1}^n \beta_0 - \beta_1 \sum_{i=1}^n x_i = 0 \text{ und } \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \beta_0 x_i - \beta_1 \sum_{i=1}^n x_i^2 = 0 \quad (11)$$
$$\Leftrightarrow \beta_0 n + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \text{ und } \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i$$

Das sich hier ergebende Gleichungssystem

$$\beta_0 n + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (12)$$
$$\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i$$

wird *System der Normalgleichungen* genannt und beschreibt die notwendige Bedingung für ein Minimum von q . Auflösen dieses Gleichungssystems nach β_0 und β_1 liefert dann die Werte $\hat{\beta}_0$ und $\hat{\beta}_1$ des Theorems.

Methode der kleinsten Quadrate

Beweis (fortgeführt)

Um dies zu sehen, halten wir zunächst fest, dass mit der ersten Gleichung des Systems der Normalgleichungen gilt

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \Leftrightarrow \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} \Leftrightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (13)$$

Einsetzen der Form von $\hat{\beta}_0$ in die zweite Gleichung des Systems der Normalgleichungen ergibt dann zunächst

$$\begin{aligned} & \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \\ \Leftrightarrow & (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \\ \Leftrightarrow & \bar{y} \sum_{i=1}^n x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \\ \Leftrightarrow & -\hat{\beta}_1 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i \\ \Leftrightarrow & \hat{\beta}_1 \left(\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right) = \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i \end{aligned} \quad (14)$$

Methode der kleinsten Quadrate

Beweis (fortgeführt)

Wir halten nun zunächst fest, dass gilt

$$\begin{aligned}\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \bar{x} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \bar{x} \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n \left(x_i^2 - 2\bar{x}x_i + \bar{x}^2 \right) \\ &= \sum_{i=1}^n \left(x_i - \bar{x} \right)^2.\end{aligned}\tag{15}$$

Methode der kleinsten Quadrate

Beweis (fortgeführt)

Weiterhin halten wir zunächst fest, dass gilt

$$\begin{aligned}\sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i - n\bar{y}\bar{x} + n\bar{y}\bar{x} \\ &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i - \sum_{i=1}^n y_i \bar{x} + \sum_{i=1}^n \bar{y} \bar{x} \\ &= \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \bar{x} - \sum_{i=1}^n \bar{y} x_i + \sum_{i=1}^n \bar{y} \bar{x} \\ &= \sum_{i=1}^n (y_i x_i - y_i \bar{x} - \bar{y} x_i + \bar{y} \bar{x}) \\ &= \sum_{i=1}^n (y_i - \bar{y}) (x_i - \bar{x}).\end{aligned}\tag{16}$$

Beweis (fortgeführt)

In der Fortsetzung von (14) ergibt sich dann

$$\begin{aligned}\hat{\beta}_1 \left(\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right) &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i \\ \Leftrightarrow \hat{\beta}_1 \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) &= \sum_{i=1}^n (y_i - \bar{y}) (x_i - \bar{x}) \\ \Leftrightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^n (y_i - \bar{y}) (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \Leftrightarrow \hat{\beta}_1 &= \frac{c_{xy}}{s_x^2}.\end{aligned}\tag{17}$$

□

Beispieldatensatz Analyse

```
# Einlesen des Beispieldatensatzes
fname      = file.path(getwd(), "1_Daten", "1_Regression.csv")
D          = read.table(fname, sep = ",", header = TRUE)

# Stichprobenstatistiken
x_bar      = mean(D$x_i)           # Stichprobenmittel der x_i-Werte
y_bar      = mean(D$y_i)           # Stichprobenmittel der y_i-Werte
s2x        = var(D$x_i)            # Stichprobenvarianz der x_i-Werte
cxy        = cov(D$x_i, D$y_i)     # Stichprobenkovarianz der (x_i, y_i)-Werte

# Ausgleichsgeradenparameter
beta_1_hat = cxy/s2x               # \hat{\beta}_1, Steigungsparameter
beta_0_hat = y_bar - beta_1_hat*x_bar # \hat{\beta}_0, Offset Parameter

# Ausgabe
cat("beta_0_hat:", beta_0_hat,
    "\nbeta_1_hat:", beta_1_hat)
```

```
> beta_0_hat: -6.19
```

```
> beta_1_hat: 1.66
```

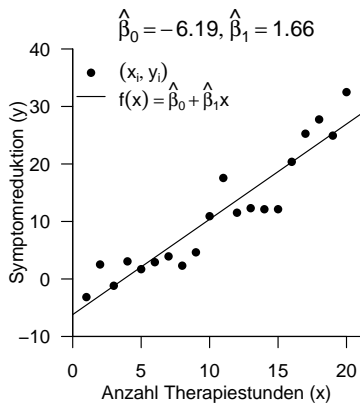
Beispieldatensatz Visualisierung

```
# Datenwerte
plot(
  D$x_i,
  D$y_i,
  pch      = 16,
  xlab     = "Anzahl Therapiestunden (x)",
  ylab     = "Symptomreduktion (y)",
  xlim     = c(0,21),
  ylim     = c(-10, 40),
  main     = TeX("\\hat{\\beta}_0 = -6.19, \\hat{\\beta}_1 = 1.66$"))

# Ausgleichsgerade
abline(
  coef     = c(beta_0_hat, beta_1_hat),
  lty      = 1,
  col      = "black")

# Legende
legend(
  "topleft",
  c(TeX("$x_i, y_i$"), TeX("$f(x) = \\hat{\\beta}_0 + \\hat{\\beta}_1 x$")),
  lty      = c(0,1),
  pch      = c(16, NA),
  bty      = "n")
```


Beispieldatensatz Visualisierung



Definition (Ausgleichspolynom)

Für $\beta := (\beta_0, \dots, \beta_k)^T \in \mathbb{R}^{k+1}$ heißt die Polynomfunktion k ten Grades

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \sum_{i=0}^k \beta_i x^i, \quad (18)$$

für die für eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ die Funktion

$$q : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^n (y_i - f_\beta(x_i))^2 = \sum_{i=1}^n \left(y_i - \sum_{i=0}^k \beta_i x^i \right)^2 \quad (19)$$

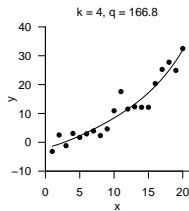
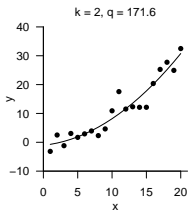
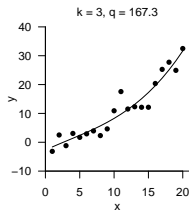
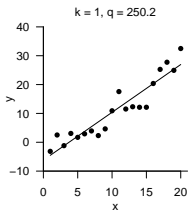
der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$ ihr Minimum annimmt, das *Ausgleichspolynom k ten Grades* für die Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\}$.

Bemerkungen

- Wir nehmen hier ohne Beweis an, dass das Minimum von q eindeutig ist.
- Die Ausgleichsgerade ist das Ausgleichspolynom ersten Grades.
- Die Parameterwerte $\hat{\beta}_0, \dots, \hat{\beta}_k$ für die q bei gegebener Wertemenge ihr Minimum annehmen werden an späterer Stelle im Rahmen der Theorie des Allgemeinen Linearen Modells ALMs bestimmt werden.

Methode der kleinsten Quadrate

Beispieldatensatz Ausgleichspolynome 1ten bis 4ten Grades



$$\bullet (x_i, y_i) \quad - \quad f_{\hat{\beta}}(x) = \sum_{i=0}^k \hat{\beta}_i x^i$$

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Motivation

Eine Ausgleichsgerade erlaubt Aussagen über unbeobachtete y Werte für x Werte. Der Wert von $q(\hat{\beta})$ quantifiziert die Güte der Ausgleichsgeradenpassung. Eine Ausgleichsgerade erlaubt allerdings nur implizite Aussagen über die mit der Anpassung verbundene Unsicherheit.

In der einfachen linearen Regression wird die Idee einer Ausgleichsgerade um eine probabilistische Komponente (normalverteilte Fehlervariable) erweitert, um quantitative Aussagen über die mit einer Ausgleichsgeradenanpassung verbundene Unsicherheit machen zu können. Weiterhin erlaubt die einfache lineare Regression, einen Hypothesentest-basierten Zugang zur Einschätzung der angepassten Parameterwerte $\hat{\beta}_0$ und $\hat{\beta}_1$.

Wir betrachten hier zunächst nur das probabilistische Modell der einfachen linearen Regression sowie die auf ihm basierende Maximum Likelihood Schätzung der Parameter β_0 und β_1 . Die Bewertung von Parameterschätzerunsicherheit sowie parameterzentrierte Hypothesentests behandeln wir an späterer Stelle zunächst im Allgemeinen.

Definition (Generatives Modell der einfachen linearen Regression)

Für $i = 1, \dots, n$ sei

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (20)$$

wobei

- $x_i \in \mathbb{R}$ fest vorgegebene sogenannte *Prädiktorwerte* oder *Regressorwerte* sind,
- $\beta_0, \beta_1 \in \mathbb{R}$ wahre, aber unbekannte, Parameterwerte sind und
- $\varepsilon_i \sim N(0, \sigma^2)$ unabhängige und identisch normalverteilte nicht-beobachtbare Zufallsvariablen mit wahren, aber unbekanntem, Parameter $\sigma^2 > 0$ sind.

Dann heißt (20) *Generatives Modell der einfachen linearen Regression*.

Bemerkungen

- Das Modell der einfachen linearen Regression hat drei Parameter, $\beta_0, \beta_1 \in \mathbb{R}$ und $\sigma^2 > 0$.

Theorem (Normalverteilungsmodell der einfachen linearen Regression)

Das generative Modell der einfachen linearen Regression

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \quad (21)$$

lässt sich äquivalent in der Form

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \text{ u.v. f\"ur } i = 1, \dots, n \quad (22)$$

schreiben.

Bemerkungen

- Wir bezeichnen

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \text{ u.v. f\"ur } i = 1, \dots, n \quad (23)$$

als *Normalverteilungsmodell der einfachen linearen Regression*.

Einfache lineare Regression

Beweis

Wir zeigen die Äquivalenz für ein i , die Unabhängigkeit der Y_i zeigen wir an späterer Stelle im Rahmen des Allgemeinen Linearen Modells. Die Äquivalenz beider Modellformen für ein i folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen (cf. (8) Transformationen der Normalverteilung). Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$, dass

$$Y_i = f(\varepsilon_i) \text{ mit } f: \mathbb{R} \rightarrow \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + (\beta_0 + \beta_1 x_i) \quad (24)$$

Mit dem WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

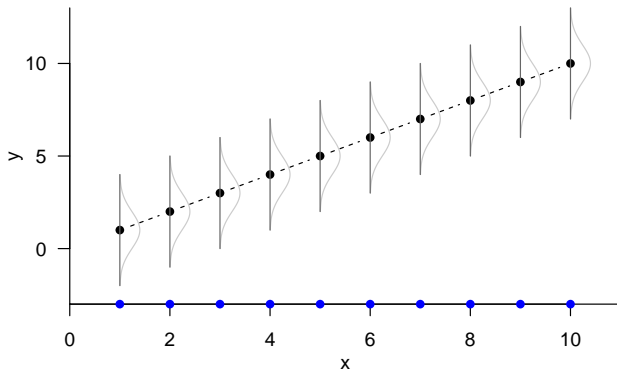
$$\begin{aligned} p_{Y_i}(y_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{1} \right) \\ &= N \left(x_i - \beta_0 - \beta_1 x_i; 0, \sigma^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \beta_0 - \beta_1 x_i - 0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - (\beta_0 + \beta_1 x_i))^2 \right) \\ &= N \left(x_i; \beta_0 + \beta_1 x_i, \sigma^2 \right), \end{aligned} \quad (25)$$

also

$$Y_i \sim N \left(\beta_0 + \beta_1 x_i, \sigma^2 \right). \quad (26)$$

Einfache lineare Regression

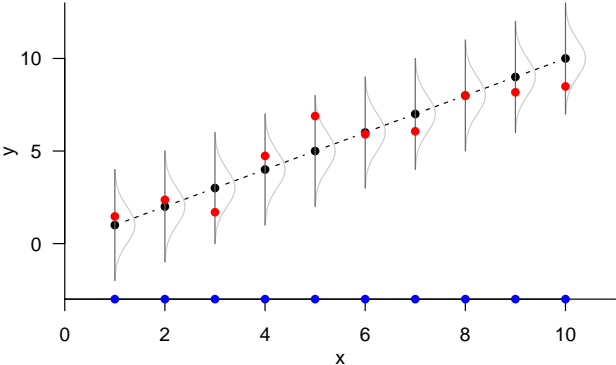
Modell der einfachen linearen Regression



• x_i • $\beta_0 + \beta_1 x_i$ für $\beta_0 := 0, \beta_1 := 1$ — $N(y_i; \beta_0 + \beta_1 x_i, \sigma^2)$ für $\sigma^2 := 1$.

Einfache lineare Regression

Realisierung des Modells der einfachen linearen Regression



• x_i • $\beta_0 + \beta_1 x_i$ für $\beta_0 := 0, \beta_1 := 1$ — $N(y_i; \beta_0 + \beta_1 x_i, \sigma^2)$ für $\sigma^2 := 1$ • (x_i, y_i)

Theorem (Maximum Likelihood Schätzung)

Es sei

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \quad (27)$$

das Modell der einfachen linearen Regression. Dann sind Maximum Likelihood Schätzer der Modellparameter β_0, β_1 und σ^2 gegeben durch

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2}, \quad \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und} \quad \hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)^2. \quad (28)$$

Bemerkungen

- Wir verzichten hier aus Gründen der Übersichtlichkeit auf die ^{ML} Superskripte.
- Die ML Schätzer für β_0 und β_1 sind offenbar mit den Ausgleichsgeradenparametern identisch.

Einfache lineare Regression

Beweis

Wir zeigen zunächst, dass die Ausgleichsgeradenparameter $\hat{\beta}_0$ und $\hat{\beta}_1$ den entsprechenden ML Schätzern gleichen. Dazu halten wir zunächst fest, dass aufgrund der Unabhängigkeit der Y_1, \dots, Y_n die Likelihood-Funktion des Modells der einfachen linearen Regression bezüglich β_0 und β_1 die Form

$$\begin{aligned} L : \mathbb{R}^2 \rightarrow \mathbb{R}_{>0}, (\beta_0, \beta_1) &\mapsto L(\beta_0, \beta_1) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2\right). \end{aligned} \quad (29)$$

Weil für die Exponentialfunktion gilt, dass für $a < b \leq 0$ gilt, dass $\exp(a) < \exp(b)$ wird der Exponentialterm dieser Likelihood-Funktion maximal, wenn der Term

$$q := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \geq 0 \quad (30)$$

minimal und damit $-q$ maximal wird. Im Rahmen des Beweises der Ausgleichsgeradenform haben wir aber schon gezeigt, dass der Term (30) für

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad (31)$$

minimal wird, und damit $\hat{\beta}_1$ und $\hat{\beta}_0$ die Likelihood-Funktion maximieren.

Einfache lineare Regression

Beweis (fortgeführt)

In einem zweiten Schritt betrachten wir nun die Likelihood-Funktion des Modells der einfachen linearen Regression bezüglich σ^2 an der Stelle von $\hat{\beta}_0$ und $\hat{\beta}_1$. Wir erhalten die Likelihood-Funktion

$$L : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, \sigma^2 \mapsto L(\sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2\right) \quad (32)$$

und die entsprechende Log-Likelihood-Funktion

$$\ell : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto \ell(\sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \quad (33)$$

In Analogie zu der Herleitung des ML Schätzers für σ^2 im Normalverteilungsmodell (cf. (10) Parameterschätzung) ergibt sich unter Beachtung von

$$\hat{\mu}_n^{\text{ML}} = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (34)$$

dann hier

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2. \quad (35)$$

Beispieldatensatz Parameterschätzung

```
# Einlesen des Beispieldatensatzes
fname = file.path(getwd(), "1_Daten", "1_Regression.csv")
D = read.table(fname, sep = ",", header = TRUE)

# Stichprobenstatistiken
n = length(D$y_i) # Anzahl Datenpunkte
x_bar = mean(D$x_i) # Stichprobenmittel der x_i-Werte
y_bar = mean(D$y_i) # Stichprobenmittel der y_i-Werte
s2x = var(D$x_i) # Stichprobenvarianz der x_i-Werte
cxy = cov(D$x_i, D$y_i) # Stichprobenkovarianz der (x_i, y_i)-Werte

# Parameterschätzer
beta_1_hat = cxy/s2x # \hat{\beta}_1, Steigungsparameter
beta_0_hat = y_bar - beta_1_hat*x_bar # \hat{\beta}_0, Offset Parameter
sigsqr_hat = (1/n)*sum((D$y_i-(beta_0_hat+beta_1_hat*D$x_i))^2) # Varianzparameter

# Ausgabe
cat("beta_0_hat:" , beta_0_hat,
    "\nbeta_1_hat:", beta_1_hat,
    "\nsigsqr_hat:", sqrt(sigsqr_hat))

> beta_0_hat: -6.19
> beta_1_hat: 1.66
> sigsqr_hat: 3.54
```

Beispieldatensatz Analyse mit `lm()`

```
# Einlesen des Beispieldatensatzes
```

```
library(car)
```

```
> Lade nötiges Paket: carData
```

```
fname = file.path(getwd(), "1_Daten", "1_Regression.csv")
```

```
D = read.table(fname, sep = ",", header = TRUE)
```

```
# Analyse mit lm()
```

```
model = lm(formula = D$y_i ~ D$x_i, data = D)
```

```
print(model)
```

```
>
```

```
> Call:
```

```
> lm(formula = D$y_i ~ D$x_i, data = D)
```

```
>
```

```
> Coefficients:
```

```
> (Intercept)      D$x_i
```

```
>      -6.19         1.66
```

Methode der kleinsten Quadrate

Einfache lineare Regression

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die funktionale Form eine linear-affinen Funktion an.
2. Erläutern Sie die Bedeutung der Parameter einer linear-affinen Funktion.
3. Definieren Sie den Begriff der Ausgleichsgerade.
4. Erläutern Sie die intuitive Bedeutung der Funktion der quadrierten vertikalen Abweichungen.
5. Geben Sie das Theorem zur Ausgleichsgerade wieder.
6. Skizzieren Sie den Beweis des Theorems zur Ausgleichsgeraden.
7. Definieren Sie den Begriff des Ausgleichspolynoms.
8. Erläutern Sie die Motivation des einfachen linearen Regressionsmodells in bezug auf die Ausgleichsgerade.
9. Definieren Sie das Generative Modell der einfachen linearen Regression.
10. Geben Sie das Theorem zum Normalverteilungsmodell der einfachen linearen Regression wieder.
11. Skizzieren das Modell der einfachen linearen Regression per Hand.
12. Skizzieren Sie eine Realisierung des Modells der einfachen linearen Regression per Hand.
13. Geben Sie das Theorem zur ML-Schätzung der Parameter der einfachen linearen Regression an.
14. Skizzieren Sie den Beweis des Theorems zur ML-Schätzung der Parameter der einfachen linearen Regression.
15. Bestimmen Sie mithilfe eines R Skripts die ML Parameterschätzer des einfachen linearen Regressionsmodells für den bereitgestellten Beispieldatensatz und visualisieren die entsprechende Regressionsgerade wie für die Ausgleichsgerade gezeigt. Geben Sie weiterhin die Bedeutung der geschätzten Parameterwerte $\hat{\beta}_0$ und $\hat{\beta}_1$ an.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(2) Korrelation

Grundlagen

Korrelation und Bestimmtheitsmaß

Korrelation und linear-affine Abhängigkeit

Korrelation und Regression

Bedingte Korrelation und Partielle Korrelation

Selbstkontrollfragen

Grundlagen

Korrelation und Bestimmtheitsmaß

Korrelation und linear-affine Abhängigkeit

Korrelation und Regression

Bedingte Korrelation und Partielle Korrelation

Selbstkontrollfragen

Anwendungsszenario

Psychotherapie



Mehr Therapiestunden

⇒ Höhere Wirksamkeit?

Unabhängige Variable

- Anzahl Therapiestunden

Abhängige Variable

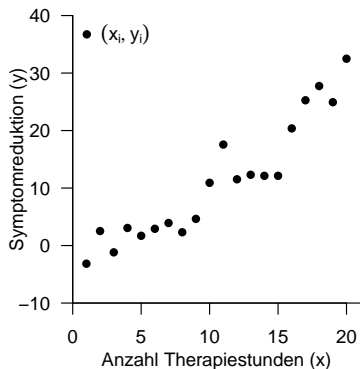
- Symptomreduktion

Beispieldatensatz

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

y_i	x_i
-3.15	1
2.52	2
-1.18	3
3.06	4
1.70	5
2.91	6
3.92	7
2.31	8
4.63	9
10.91	10
17.56	11
11.52	12
12.31	13
12.12	14
12.13	15
20.37	16
25.26	17
27.75	18
24.93	19
32.49	20

Beispieldatensatz



Wie stark hängen Anzahl Therapiestunden und Symptomreduktion zusammen?

Definition (Korrelation)

Die *Korrelation* zweier Zufallsvariablen X und Y ist definiert als

$$\rho(X, Y) := \frac{\mathbb{C}(X, Y)}{\mathbb{S}(X)\mathbb{S}(Y)} \quad (1)$$

wobei $\mathbb{C}(X, Y)$ die Kovarianz von X und Y und $\mathbb{S}(X)$ und $\mathbb{S}(Y)$ die Standardabweichungen von X und Y , respektive, bezeichnen.

Bemerkungen

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

Definition (Stichprobenkorrelation)

$\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}$ sei eine Wertemenge. Weiterhin seien:

- Die Stichprobenmittel der x_i und y_i definiert als

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

- Die Stichprobenstandardabweichungen x_i und y_i definiert als

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

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

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

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

$$r_{xy} := \frac{c_{xy}}{s_x s_y} \quad (5)$$

und wird auch *Stichprobenkorrelationskoeffizient* genannt.

Beispiel

```
# Laden des Beispieldatensatzes
fname = file.path(getwd(), "2_Daten", "2_Korrelation_Beispieldatensatz.csv") # Dateipfad
D      = read.table(fname, sep = ",", header = TRUE)                       # Laden als Dataframe
x_i    = D$x_i                                                             # x_i Werte
y_i    = D$y_i                                                             # 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
```

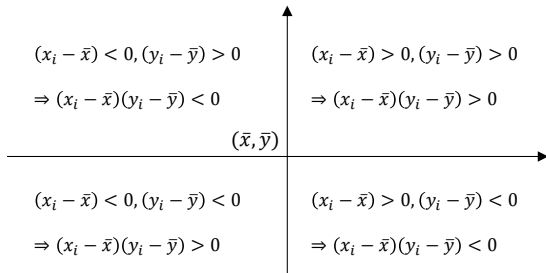
```
> [1] 0.938
```

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

```
> [1] 0.938
```

⇒ Anzahl Therapiestunden und Symptomreduktion sind hochkorreliert.

Mechanik der Kovariationsterme

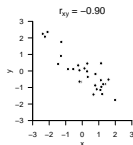
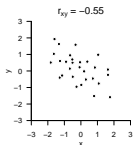
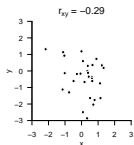
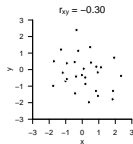
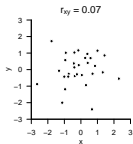
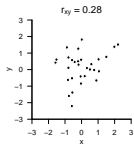
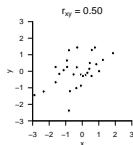
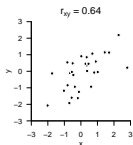
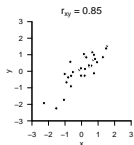


Häufige richtungsgleiche Abweichung der x_i und y_i von ihren Mittelwerten \Rightarrow Positive Korrelation

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

Keine häufigen richtungsgleichen oder -entgegengesetzten Abweichungen \Rightarrow Keine Korrelation

Beispiele



Theorem (Stichprobenkorrelation bei linear-affinen Transformationen)

Für eine Wertemenge $\{(x_i, y_i)\}_{i=1, \dots, n} \subset \mathbb{R}^2$ sei $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1, \dots, n} \subset \mathbb{R}^2$ eine linear-affin transformierte Wertemenge mit

$$(\tilde{x}_i, \tilde{y}_i) = (a_x x_i + b_x, a_y y_i + b_y), a_x, a_y \neq 0. \quad (6)$$

Dann gilt

$$|r_{\tilde{x}\tilde{y}}| = |r_{xy}|. \quad (7)$$

Bemerkungen

- Der Betrag der Stichprobenkorrelation ändert sich bei linear-affiner Datentransformation nicht.
- Man sagt, dass die Stichprobenkorrelation im Gegensatz zur Stichprobenkovarianz *maßstabsunabhängig* ist.

Grundlagen

Beweis

Es gilt

$$\begin{aligned} r_{\hat{x}\hat{y}} &:= \frac{\frac{1}{n-1} \sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n \hat{x}_i - \bar{\hat{x}} \right)^2} \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n \hat{y}_i - \bar{\hat{y}} \right)^2}} \\ &= \frac{\sum_{i=1}^n (a_x x_i + b_x - (a_x \bar{x} + b_x))(a_y y_i + b_y - (a_y \bar{y} + b_y))}{\sqrt{\sum_{i=1}^n (a_x x_i + b_x - (a_x \bar{x} + b_x))^2} \sqrt{\sum_{i=1}^n (a_y y_i + b_y - (a_y \bar{y} + b_y))^2}} \\ &= \frac{a_x a_y \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{a_x^2 \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{a_y^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{a_x a_y}{|a_x| |a_y|} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{a_x a_y}{|a_x| |a_y|} \frac{c_{xy}}{s_x s_y} \\ &= \frac{a_x a_y}{|a_x| |a_y|} r_{xy}. \end{aligned} \tag{8}$$

Also folgt, durch Durchspielen aller möglichen Vorzeichenfälle, dass

$$|r_{\hat{x}\hat{y}}| = |r_{xy}|. \tag{9}$$

□

Grundlagen

Korrelation und Bestimmtheitsmaß

Korrelation und linear-affine Abhängigkeit

Korrelation und Regression

Bedingte Korrelation und Partielle Korrelation

Selbstkontrollfragen

Überblick

Das sogenannte Bestimmtheitsmaß R^2 ist eine beliebte Statistik.

Numerisch ist R^2 das Quadrat des Stichprobenkorrelationskoeffizienten.

Ist die Stichprobenkorrelation $r_{xy} = 0.5$, dann ist $R^2 = 0.25$, ist $r_{xy} = -0.5$, dann ist $R^2 = 0.25$.

⇒ R^2 enthält also weniger Information über die Rohdaten als r_{xy} , da das Vorzeichen wegfällt.

⇒ *Perse* ist die Angabe von R^2 anstelle von r_{xy} im Kontext der Korrelation zweier Variablen wenig sinnvoll.

Ein tieferes Verständnis von R^2 erlaubt jedoch

- (1) Einen Einstieg in das Konzept von Quadratsummenzerlegungen, einem wichtigen ALM Evaluationsprinzip.
- (2) Einen Einstieg in das Verständnis der Zusammenhänge von Ausgleichsgerade und Stichprobenkorrelation.
- (3) Einen ersten Einblick in die Tatsache, dass Korrelationen (nur) linear-affine Zusammenhänge quantifizieren.

Definition (Erklärte Werte und Residuen einer Ausgleichsgerade)

Gegeben seien eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ und die zu dieser Wertemenge gehörende Ausgleichsgerade

$$f_{\hat{\beta}} : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_{\hat{\beta}}(x) := \hat{\beta}_0 + \hat{\beta}_1 x \quad (10)$$

Dann werden für $i = 1, \dots, n$

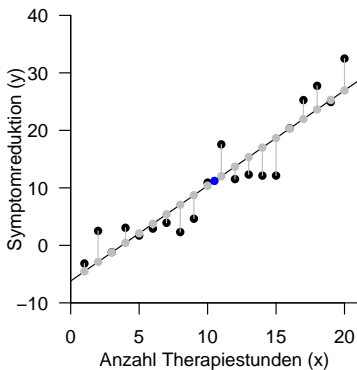
$$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (11)$$

die durch die Ausgleichsgerade *erklärten Werte* genannt und

$$\hat{\varepsilon}_i := y_i - \hat{y}_i \quad (12)$$

die *Residuen* der Ausgleichsgerade genannt.

Erklärte Werte und Residuen



• (x_i, y_i) • (\bar{x}, \bar{y}) — $f_{\hat{\beta}}(x)$ • \hat{y}_i — $\hat{\varepsilon}_i$ $i = 1, \dots, n$

Theorem (Quadratsummenzerlegung bei Ausgleichsgerade)

Für eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ und ihre zugehörige Ausgleichsgerade $f_{\hat{\beta}}$ seien für

$$\bar{y} := \frac{1}{n} \sum_{i=1}^n y_i \text{ und } \hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \text{ für } i = 1, \dots, n \quad (13)$$

das Stichprobenmittel der y -Werte und die durch die Ausgleichsgerade erklärten Werte, respektive. Weiterhin seien

$$\text{SQT} := \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{die Total Sum of Squares}$$

$$\text{SQE} := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{die Explained Sum of Squares}$$

$$\text{SQR} := \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{die Residual Sum of Squares}$$

Dann gilt

$$\text{SQT} = \text{SQE} + \text{SQR} \quad (14)$$

Bemerkungen

- SQT repräsentiert die Gesamtstreuung der y_i -Werte um ihren Mittelwert \bar{y} .
- SQE repräsentiert die Streuung der erklärten Werte \hat{y}_i um ihren Mittelwert
 - ⇒ Große Werte von SQE repräsentieren eine große absolute Steigung der y_i mit den x_i
 - ⇒ Kleine Werte von SQE repräsentieren eine kleine absolute Steigung der y_i mit den x_i
- SQE ist also ein Maß für die Stärke des linearen Zusammenhangs der x - und y -Werte
- SQR ist die Summe der quadrierten Residuen, es gilt

$$\text{SQR} := \sum_{i=1}^n (y_i - \hat{y}_i)^2 := \sum_{i=1}^n \hat{\epsilon}_i^2 \quad (15)$$

- ⇒ Große Werte von SQR repräsentieren große Abweichungen der erklärten von den beobachteten y -Werten
- ⇒ Kleine Werte von SQR repräsentieren geringe Abweichungen der erklärten von den beobachteten y -Werten
- SQR ist also ein Maß für die Güte der Beschreibung der Datenmenge durch die Ausgleichsgerade.

Beweis

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n ((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2 \\ &= \sum_{i=1}^n \left((y_i - \hat{y}_i)^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \text{SQE} + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \text{SQR} \\ &= \text{SQE} + \text{SQR} \end{aligned} \tag{16}$$

Beweis (fortgeführt)

Dabei ergibt sich die letzte Gleichung mit

$$\bar{\hat{y}} := \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y} \quad (17)$$

und damit auch

$$\bar{\hat{y}} = \bar{y} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \Leftrightarrow \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i \Leftrightarrow \bar{y} \sum_{i=1}^n \hat{y}_i = \bar{y} \sum_{i=1}^n y_i \quad (18)$$

sowie

$$\bar{\hat{y}} = \bar{y} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \Leftrightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i \Leftrightarrow \sum_{i=1}^n y_i \hat{y}_i = \sum_{i=1}^n \hat{y}_i \hat{y}_i \quad (19)$$

aus

Beweis (fortgeführt)

$$\begin{aligned}\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n (y_i \hat{y}_i - y_i \bar{y} - \hat{y}_i \hat{y}_i + \hat{y}_i \bar{y}) \\ &= \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n y_i \bar{y} - \sum_{i=1}^n \hat{y}_i \hat{y}_i + \sum_{i=1}^n \hat{y}_i \bar{y} \\ &= \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n \hat{y}_i \hat{y}_i + \bar{y} \sum_{i=1}^n \hat{y}_i - \bar{y} \sum_{i=1}^n y_i \\ &= 0 + 0 \\ &= 0\end{aligned}\tag{20}$$

□

Definition (Bestimmtheitsmaß R^2)

Für eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ und ihre zugehörige Ausgleichsgerade $f_{\hat{\beta}}$ sowie die zugehörigen Explained Sum of Squares SQE und Total Sum of Squares SQT heißt

$$R^2 := \frac{\text{SQE}}{\text{SQT}} \quad (21)$$

Bestimmtheitsmaß oder Determinationskoeffizient.

Theorem (Stichprobenkorrelation und Bestimmtheitsmaß)

Für eine Wertemenge $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ sei R^2 das Bestimmtheitsmaß und r_{xy} sei die Stichprobenkorrelation. Dann gilt

$$R^2 = r_{xy}^2. \quad (22)$$

Bemerkungen

- Mit $-1 \leq r_{xy} \leq 1$ folgt aus dem Theorem direkt, dass $0 \leq R^2 \leq 1$.
- Es gilt $R^2 = 0$ genau dann, wenn $SQE = 0$ ist
 - ⇒ Für $R^2 = 0$ ist die erklärte Streuung der Daten durch die Ausgleichsgerade gleich null.
 - ⇒ $R^2 = 0$ beschreibt also den Fall einer denkbar schlechten Erklärung der Daten durch die Ausgleichsgerade.
- Es gilt $R^2 = 1$ genau dann, wenn $SQE = SQT$ ist.
 - ⇒ Für $R^2 = 1$ ist also die Gesamtstreuung gleich der durch die Ausgleichsgerade erklärten Streuung.
 - ⇒ $R^2 = 1$ beschreibt also den Fall das sämtliche Datenvariabilität durch die Ausgleichsgerade erklärt wird.
- Man sagt, dass " R^2 die durch die Ausgleichsgerade erklärte Varianz an der Gesamtdatenvarianz" ist.

Korrelation und Bestimmtheitsmaß

Beweis

Wir halten zunächst fest, dass mit

$$\bar{\hat{y}} := \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y} \quad (23)$$

folgt, dass

$$\begin{aligned} \text{SQE} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \\ &= \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})^2 \\ &= \sum_{i=1}^n (\hat{\beta}_1 (x_i - \bar{x}))^2 \\ &= \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned} \quad (24)$$

Korrelation und Bestimmtheitsmaß

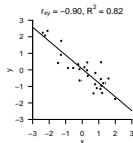
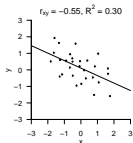
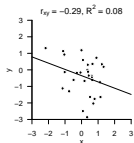
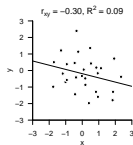
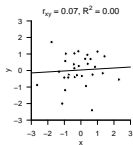
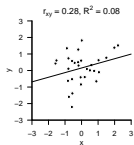
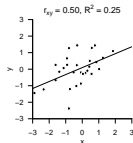
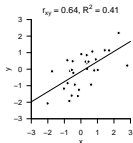
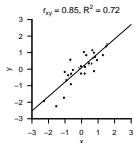
Beweis

Damit ergibt sich dann

$$\begin{aligned} R^2 &= \frac{\text{SQE}}{\text{SQT}} \\ &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \hat{\beta}_1^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{c_{xy}^2 \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}{s_x^4 \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{c_{xy}^2}{s_x^4} \frac{s_x^2}{s_y^2} \\ &= \frac{c_{xy}^2}{s_x^2 s_y^2} \\ &= \left(\frac{c_{xy}}{s_x s_y} \right)^2 \\ &= r_{xy}^2. \end{aligned} \tag{25}$$

□

Beispiele



Grundlagen

Korrelation und Bestimmtheitsmaß

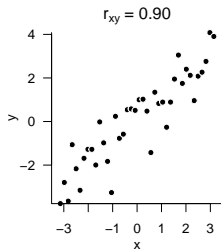
Korrelation und linear-affine Abhängigkeit

Korrelation und Regression

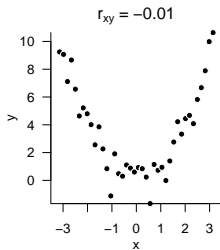
Bedingte Korrelation und Partielle Korrelation

Selbstkontrollfragen

Funktionale Abhängigkeiten und Stichprobenkorrelation

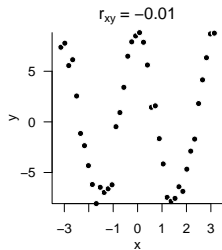


$$y_i = x_i + \varepsilon_i$$



$$y_i = x_i^2 + \varepsilon_i$$

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



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

Theorem (Korrelation und linear-affine Abhängigkeit)

X und Y seien zwei Zufallsvariablen mit positiver Varianz. Dann besteht genau dann eine linear-affine Abhängigkeit der Form

$$Y = \beta_0 + \beta_1 X \text{ mit } \beta_0, \beta_1 \in \mathbb{R} \quad (26)$$

zwischen X und Y , wenn

$$\rho(X, Y) = 1 \text{ oder } \rho(X, Y) = -1 \quad (27)$$

gilt.

Bemerkungen

- Die linear-affine Abhängigkeit $Y = \beta_0 + \beta_1 X$ impliziert eine linear-affine Abhängigkeit $X = \tilde{\beta}_0 + \tilde{\beta}_1 Y$, denn

$$Y = \beta_0 + \beta_1 X \Leftrightarrow -\beta_0 + Y = \beta_1 X \Leftrightarrow X = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} Y \Leftrightarrow X = \tilde{\beta}_0 + \tilde{\beta}_1 Y \quad (28)$$

mit

$$\tilde{\beta}_0 = -\frac{\beta_0}{\beta_1} \text{ und } \tilde{\beta}_1 = \frac{1}{\beta_1}. \quad (29)$$

Korrelation und linear-affine Abhängigkeit

Beweis

Wir beschränken uns auf den Beweis der Aussage, dass aus $Y = \beta_0 + \beta_1 X$ folgt, dass $\rho(X, Y) = \pm 1$ ist. Dazu halten wir zunächst fest, dass mit den Theoremen zu den Eigenschaften von Erwartungswert und Varianz gilt, dass

$$\mathbb{E}(Y) = \beta_0 + \beta_1 \mathbb{E}(X) \text{ und } \mathbb{V}(Y) = \beta_1^2 \mathbb{V}(X). \quad (30)$$

Wegen $\mathbb{V}(X) > 0$ und $\mathbb{V}(Y) > 0$ gilt damit $\beta_1 \neq 0$. Es folgt dann

$$\beta_1 > 0 \Rightarrow \mathbb{S}(Y) = \beta_1 \mathbb{S}(X) > 0 \text{ und } \beta_1 < 0 \Rightarrow \mathbb{S}(Y) = -\beta_1 \mathbb{S}(X) > 0. \quad (31)$$

Weiterhin gilt

$$\begin{aligned} Y - \mathbb{E}(Y) &= \beta_0 + \beta_1 X - \mathbb{E}(Y) \\ &= \beta_0 + \beta_1 X - \beta_0 - \beta_1 \mathbb{E}(X) \\ &= \beta_1 X - \beta_1 \mathbb{E}(X) \\ &= \beta_1 (X - \mathbb{E}(X)). \end{aligned} \quad (32)$$

Für die Kovarianz von X und Y ergibt sich also

$$\begin{aligned} \mathbb{C}(X, Y) &= \mathbb{E}((Y - \mathbb{E}(Y))(X - \mathbb{E}(X))) \\ &= \mathbb{E}(\beta_1 (X - \mathbb{E}(X))(X - \mathbb{E}(X))) \\ &= \beta_1 \mathbb{E}((X - \mathbb{E}(X))^2) \\ &= \beta_1 \mathbb{V}(X). \end{aligned} \quad (33)$$

Damit ergibt für die Korrelation von X und Y

$$\rho(X, Y) = \frac{\mathbb{C}(X, Y)}{\mathbb{S}(X)\mathbb{S}(Y)} = \pm \frac{\beta_1 \mathbb{V}(X)}{\mathbb{S}(X)\beta_1 \mathbb{S}(X)} = \pm \frac{\beta_1 \mathbb{V}(X)}{\beta_1 \mathbb{V}(X)} = \pm 1. \quad (34)$$

Grundlagen

Korrelation und Bestimmtheitsmaß

Korrelation und linear-affine Abhängigkeit

Korrelation und Regression

Bedingte Korrelation und Partielle Korrelation

Selbstkontrollfragen

Überblick

Der fundamentale Unterschied zwischen "Korrelation" und "Regression" ist, dass

- bei Korrelation sowohl die UV (die x 's) als auch die AV (die y 's) als Zufallsvariablen modelliert werden,
- bei Regression dagegen lediglich die AV als Zufallsvariable modelliert wird und die UV als vorgegeben gilt.

Dieser Tatsache unbenommen, kann man auf gegebene Daten prinzipiell natürlich sowohl "Korrelation" als auch "Regression" anwenden. Das Ergebnis einer Regressionsanalyse lässt sich in das Ergebnis einer Korrelationsanalyse umrechnen. Die zusätzlich Durchführung einer Korrelationsanalyse bei durchgeführter Regressionsanalyse erzeugt kein mehr an Information oder Verständnis über den Zusammenhang von UV und AV.

Für ein tieferes Verständnis dieser Zusammenhänge ist ein Regressionsmodell nötig, indem auch die UV eine Zufallsvariable ist. In Abgrenzung zum Modell der einfachen linearen Regression, in dem die UV keine Zufallsvariable ist, bezeichnen wir dieses Modell als *Regression*. Letztlich gerät die Terminologie hier an eine Grenze und es muss jeweils geprüft bzw. geschlossen werden, welches Modell Datenanalysten nun tatsächlich vorschwebt.

Weiterhin treffen wir die Annahme, dass sowohl die UV als auch die AV im Regressionsmodell normalverteilt sind. Diese Annahme ist nicht zwingend nötig, da Aussagen zur Regression zweier Zufallsvariablen im Wesentlichen die Erwartungswerte, Varianzen, und Kovarianzen berühren. Allerdings wird die Normalverteilungsannahme für UV und AV im Anwendungskontext häufig getroffen und bereitet didaktisch sinnvoll auf das Konzept multivariater Normalverteilungen vor, das für die ALM Theorie zentral ist.

Theorem (Bivariate Normalverteilung)

Z_1 und Z_2 seien zwei standardnormalverteilte Zufallsvariablen und $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 > 0, \sigma_2 > 0$ und $\rho \in]-1, 1[$ seien Konstanten. Weiterhin seien x und z zwei Zufallsvariablen definiert als

$$\begin{aligned}x &:= \sigma_1 Z_1 + \mu_1 \\z &:= \sigma_2 \left(\rho Z_1 + (1 - \rho^2)^{\frac{1}{2}} Z_2 \right) + \mu_2\end{aligned}\tag{35}$$

Dann hat die gemeinsame WDF von x und z die Form

$$\begin{aligned}p : \mathbb{R}^2 \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ z \end{pmatrix} \mapsto p \left(\begin{pmatrix} x \\ z \end{pmatrix} \right) &= \frac{1}{2\pi(1 - \rho^2)^{1/2}\sigma_1\sigma_2} \\&\times \exp \left(-\frac{1}{2(1 - \rho^2)} \left(\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\left(\frac{x - \mu_1}{\sigma_1} \right) \left(\frac{z - \mu_2}{\sigma_2} \right) \right) + \left(\frac{z - \mu_2}{\sigma_2} \right)^2 \right) \right)\end{aligned}\tag{36}$$

Definition (Korrelationsmatrix eines Zufallsvektors)

y sei ein n -dimensionaler Zufallsvektor. Die *Korrelationsmatrix* von y ist definiert als

$$\mathbb{R}(y) = \left(\rho(y_i, y_j) \right)_{1 \leq i, j \leq n} = \begin{pmatrix} \rho(y_1, y_1) & \rho(y_1, y_2) & \cdots & \rho(y_1, y_n) \\ \rho(y_2, y_1) & \rho(y_2, y_2) & \cdots & \rho(y_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(y_n, y_1) & \rho(y_n, y_2) & \cdots & \rho(y_n, y_n) \end{pmatrix}. \quad (37)$$

Bemerkung

- Die Korrelationsmatrix $\mathbb{R}(y)$ ist also die Matrix der Korrelationen der Komponenten von y .

Theorem (Korrelationsmatrix und Kovarianzmatrix)

y sei ein n -dimensionaler Zufallsvektor mit Kovarianzmatrix $\mathbb{C}(y)$. Weiterhin sei

$$S_y := \text{diag} \left(\sqrt{\mathbb{C}(y_1, y_1)}, \sqrt{\mathbb{C}(y_2, y_2)}, \dots, \sqrt{\mathbb{C}(y_n, y_n)} \right) \quad (38)$$

die Diagonalmatrix der Standardabweichungen der Komponenten von y . Dann gelten

$$\mathbb{R}(y) = S_y^{-1} \mathbb{C}(y) S_y^{-1} \quad (39)$$

und

$$\mathbb{C}(y) = S_y \mathbb{R}(y) S_y \quad (40)$$

Bemerkungen

- $\mathbb{C}(y)$ kann mithilfe ihrer Diagonalelemente in $\mathbb{R}(y)$ umgerechnet werden.
- $\mathbb{R}(y)$ kann mithilfe der Diagonalelemente von $\mathbb{C}(y)$ in $\mathbb{C}(y)$ umgerechnet werden.
- $\mathbb{C}(y)$ enthält also mehr Information als $\mathbb{R}(y)$.

Definition und Eigenschaften

Beweis

Wir halten zunächst fest, dass

$$S_y^{-1} = \text{diag} \left(\frac{1}{\sqrt{C(y_1, y_1)}}, \frac{1}{\sqrt{C(y_2, y_2)}}, \dots, \frac{1}{\sqrt{C(y_n, y_n)}} \right) \quad (41)$$

Es ergibt sich dann

$$\begin{aligned} S_y^{-1} C(y) S_y^{-1} &= \left(\frac{C(y_i, y_j)}{\sqrt{C(y_i, y_i)}} \right)_{1 \leq i, j \leq n} S_y^{-1} \\ &= \left(\frac{C(y_i, y_j)}{\sqrt{C(y_i, y_i)} \sqrt{C(y_j, y_j)}} \right)_{1 \leq i, j \leq n} \\ &= \left(\rho(y_i, y_j) \right)_{1 \leq i, j \leq n} \\ &=: R(y) \end{aligned} \quad (42)$$

Beweis

Analog ergibt sich

$$\begin{aligned} S_y \mathbb{R}(y) S_y &= \left(\frac{\sqrt{\mathbb{C}(y_i, y_i) \mathbb{C}(y_i, y_j)}}{\sqrt{\mathbb{C}(y_i, y_i)} \sqrt{\mathbb{C}(y_j, y_j)}} \right)_{1 \leq i, j \leq n} S_y \\ &= \left(\frac{\mathbb{C}(y_i, y_j) \sqrt{\mathbb{C}(y_j, y_j)}}{\sqrt{\mathbb{C}(y_j, y_j)}} \right)_{1 \leq i, j \leq n} \\ &= \left(\mathbb{C}(y_i, y_j) \right)_{1 \leq i, j \leq n} \\ &=: \mathbb{C}(y). \end{aligned} \tag{43}$$

Theorem (Korrelationsmatrix eines normalverteilten Zufallsvektors)

$y \sim N(\mu, \Sigma)$ sei ein multivariat normalverteilter Zufallsvektor mit Erwartungswertparameter $\mu \in \mathbb{R}^n$ und Kovarianzmatrixparameter

$$\Sigma := \left(\sigma_{ij}^2 \right)_{1 \leq i, j \leq n} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2 \end{pmatrix} \in \mathbb{R}^{n \times n} \text{ p.d.} \quad (44)$$

Dann gilt

$$\mathbb{R}(y) = \left(\frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}} \right)_{1 \leq i, j \leq n} = \begin{pmatrix} 1 & \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} & \cdots & \frac{\sigma_{1n}^2}{\sigma_{11}\sigma_{nn}} \\ \frac{\sigma_{21}^2}{\sigma_{22}\sigma_{11}} & 1 & \cdots & \frac{\sigma_{2n}^2}{\sigma_{22}\sigma_{nn}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}^2}{\sigma_{nn}\sigma_{11}} & \frac{\sigma_{n2}^2}{\sigma_{nn}\sigma_{22}} & \cdots & 1 \end{pmatrix}. \quad (45)$$

Bemerkungen

- Das Theorem folgt direkt mit dem Theorem zu Korrelationsmatrix und Kovarianzmatrix

Definition und Eigenschaften

Bemerkungen (fortgeführt)

- Umgekehrt gilt bei Definition eines Korrelationsmatrixparameters

$$R := (\rho_{ij})_{1 \leq i, j \leq n} = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{pmatrix} \quad (46)$$

mit

$$\rho_{ij} := 1 \text{ für } i = j \text{ und } \rho_{ij} = \rho_{ji} \in]-1, 1[\text{ für } i \neq j \text{ für } 1 \leq i, j \leq n \quad (47)$$

und zusätzlicher Definition von

$$S_y := \text{diag}(\sigma_{11}, \dots, \sigma_{nn}) \text{ mit } \sigma_{11}, \dots, \sigma_{nn} > 0, \quad (48)$$

dass

$$\Sigma := S_y R S_y = \begin{pmatrix} \sigma_{11}\sigma_{11} & \rho_{12}\sigma_{11}\sigma_{22} & \cdots & \rho_{1n}\sigma_{11}\sigma_{nn} \\ \rho_{21}\sigma_{22}\sigma_{11} & \sigma_{22}\sigma_{22} & \cdots & \rho_{2n}\sigma_{22}\sigma_{nn} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_{nn}\sigma_{11} & \rho_{n2}\sigma_{nn}\sigma_{22} & \cdots & \sigma_{nn}\sigma_{nn} \end{pmatrix} \quad (49)$$

einen Kovarianzmatrixparameter eines multivariaten normalverteilten Zufallsvektors definiert.

Definition (Regressionsgerade zweier Zufallsvariablen)

X und Y seien zwei Zufallsvariablen. Dann heißt

$$Y = \beta_0 + \beta_1 X \text{ mit} \quad (50)$$

mit

$$\beta_1 := \frac{\mathbb{C}(X, Y)}{\mathbb{V}(X)} \text{ und } \beta_0 := \mathbb{E}(Y) - \beta_1 \mathbb{E}(X) \quad (51)$$

die *Regressionsgerade der Zufallsvariablen X auf Y* , β_0 und β_1 heißen die zugehörigen *Regressionskoeffizienten*, und die Zufallsvariable

$$E := Y - \beta_0 - \beta_1 X \quad (52)$$

heißt die *Residualvariable*.

Bemerkungen

- X und Y sind Zufallsvariablen, β_0 und β_1 sind keine Zufallsvariablen.

Theorem (Optimalität der Regressionsgerade zweier Zufallsvariablen)

Unter allen Geraden der Form

$$Y = \beta_0 + \beta_1 X \quad (53)$$

ist die Gerade mit

$$\beta_1 := \frac{\mathbb{C}(X, Y)}{\mathbb{V}(X)} \text{ und } \beta_0 := \mathbb{E}(Y) - \beta_1 \mathbb{E}(X) \quad (54)$$

diejenige, für die

$$\tilde{q} : \mathbb{R}^2 \rightarrow \mathbb{R}, (\beta_0, \beta_1) \mapsto \tilde{q}(\beta_0, \beta_1) := \mathbb{E} \left((Y - (\beta_0 + \beta_1 X))^2 \right) \quad (55)$$

ein Minimum hat.

Korrelation und Regression

Beweis

Wir halten zunächst fest, dass

$$\begin{aligned}\bar{q}(\beta_0, \beta_1) &= \mathbb{E}(Y - \beta_0 - \beta_1 X) \\ &= \mathbb{E}(Y - \beta_1 X - \beta_0 + \beta \mathbb{E}(X) - \beta_1 \mathbb{E}(X) + \mathbb{E}(Y) - \mathbb{E}(Y)) \\ &= \mathbb{E}((Y - \mathbb{E}(Y)) - \beta_1 (X - \mathbb{E}(X)) + (\mathbb{E}(Y) - \beta_1 \mathbb{E}(X) - \beta_0))\end{aligned}\quad (56)$$

Ausmultiplizieren und Anwendung des Theorems zu den Eigenschaften des Erwartungswerts ergibt dann

$$\bar{q}(\beta_0, \beta_1) = \mathbb{V}(Y) + \beta_1^2 \mathbb{V}(X) - 2\beta_1 \mathbb{C}(X, Y) + (\mathbb{E}(Y) - \beta_1 \mathbb{E}(X) - \beta_0)^2 \quad (57)$$

Berechnen der partiellen Ableitungen von \bar{q} hinsichtlich von β_0 und β_1 ergibt dann

$$\frac{\partial}{\partial \beta_0} \bar{q}(\beta_0, \beta_1) = -2(\mathbb{E}(Y) - \beta_1 \mathbb{E}(X) - \beta_0) \quad (58)$$

und

$$\frac{\partial}{\partial \beta_1} \bar{q}(\beta_0, \beta_1) = 2\beta_1 \mathbb{V}(X) - 2\mathbb{C}(X, Y) - 2\mathbb{E}(X)(\mathbb{E}(Y) - \beta_1 \mathbb{E}(X) - \beta_0) \quad (59)$$

Nullsetzen von (58) ergibt dann als notwendige Bedingungen für ein Minimum von \bar{q}

$$\begin{aligned}\frac{\partial}{\partial \beta_0} \bar{q}(\beta_0^*, \beta_1^*) &= 0 \Leftrightarrow \mathbb{E}(Y) - \beta_1^* \mathbb{E}(X) - \beta_0^* = 0 \\ \frac{\partial}{\partial \beta_1} \bar{q}(\beta_0^*, \beta_1^*) &= 0 \Leftrightarrow 2\beta_1^* \mathbb{V}(X) - 2\mathbb{C}(X, Y) - 2\mathbb{E}(X)(\mathbb{E}(Y) - \beta_1^* \mathbb{E}(X) - \beta_0^*) = 0\end{aligned}\quad (60)$$

Die erste Gleichung impliziert dann für die zweite Gleichung, dass

$$2\beta_1^* \mathbb{V}(X) - 2\mathbb{C}(X, Y) = 0 \Leftrightarrow \beta_1^* = \frac{\mathbb{C}(X, Y)}{\mathbb{V}(X)} \quad (61)$$

Einsetzen in die erste Gleichung ergibt dann

$$\mathbb{E}(Y) - \beta_1^* \mathbb{E}(X) - \beta_0^* = 0 \Leftrightarrow \beta_0^* = \mathbb{E}(Y) - \beta_1^* \mathbb{E}(X) \quad (62)$$

Theorem (Zusammenhang von Korrelation und Regression)

X und Y seien zwei Zufallsvariablen,

$$Y = \beta_0 + \beta_1 X \text{ mit } \beta_1 := \frac{\mathbb{C}(X, Y)}{\mathbb{V}(X)} \text{ und } \beta_0 := \mathbb{E}(Y) - \tilde{\beta}_1 \mathbb{E}(X) \quad (63)$$

sei die Regressionsgerade der Zufallsvariablen Y bezüglich der Zufallsvariablen X mit den Regressionskoeffizienten β_0 und β_1 und

$$X = \tilde{\beta}_0 + \tilde{\beta}_1 Y \text{ mit } \tilde{\beta}_1 := \frac{\mathbb{C}(X, Y)}{\mathbb{V}(Y)} \text{ und } \tilde{\beta}_0 := \mathbb{E}(X) - \tilde{\beta}_1 \mathbb{E}(Y) \quad (64)$$

sei die Regressionsgerade der Zufallsvariablen X bezüglich der Zufallsvariablen Y mit den Regressionskoeffizienten $\tilde{\beta}_0$ und $\tilde{\beta}_1$. Dann gilt

$$\beta_1 \tilde{\beta}_1 = \frac{\mathbb{C}(X, Y)}{\mathbb{V}(X)} \frac{\mathbb{C}(X, Y)}{\mathbb{V}(Y)} = \frac{\mathbb{C}(X, Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} = \rho(X, Y)^2. \quad (65)$$

Bemerkungen

- $\rho(X, Y)$ kann aus den Regressionskoeffizienten von X auf Y und von Y auf X errechnet werden.

Definition (Stichprobenregressionsgerade)

$(x, Y_1), \dots, (X_n, Y_n)$ sei eine Stichprobe von zweidimensionalen Zufallsvektoren mit identischen unabhängigen Verteilungen. Weiterhin sei für $i = 1, \dots, n$

$$Y_1 = \beta_0 + \beta_1 x \text{ mit } \beta_1 := \frac{\mathbb{C}(x, Y_1)}{\mathbb{V}(x)} \text{ und } \beta_0 := \beta_1 \mathbb{E}(x) + \mathbb{E}(Y_1) \quad (66)$$

die Regressionsgerade der Zufallsvariablen Y_1 bezüglich der Zufallsvariablen x mit den Regressionskoeffizienten β_0 und β_1 . Schließlich seien

- \bar{x} und \bar{y} die Stichprobenmittel von Realisierungen der Komponenten der Stichprobe,
- s_X^2 und s_Y^2 die Stichprobenvarianzen von Realisierungen der Komponenten der Stichprobe und
- $c_{X,Y}$ die Stichprobenkovarianz von Realisierungen der Stichprobe.

Dann heißt für $x \in \mathbb{R}$

$$y = b_0 + b_1 x \text{ mit } b_1 := \frac{c_{X,Y}}{s_X^2} \text{ und } b_0 := \bar{y} - b_1 \bar{x} \quad (67)$$

die Regressionsgerade der y -Werte bezüglich der x_i Werte in der Stichprobe.

Korrelation und Regression

Simulation einer Regressionsgerade

```
library(MASS) # multivariate Normalverteilungen

# Modellformulierung
n = 1e2 # Anzahl an Stichprobenvektoren
C_XY = 1 # Kovarianz von X und Y
EX = 2 # Erwartungswert von X
EY = 1 # Erwartungswert von Y
VX = 2 # Varianz von X
VY = 2 # Varianz von Y
beta_1 = C_XY/VX # Regressionskoeffizient
beta_0 = -beta_1*EX + EY # Regressionskoeffizient

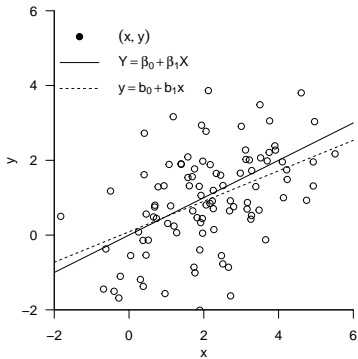
# Realisierungsgeneration
mu = c(EX, EY) # Erwartungswertparameter
Sigma = matrix(c(VX, C_XY, C_XY, VY), nrow = 2) # Kovarianzmatrixparameter
xy = mvrnorm(n, mu, Sigma)

# Stichprobenstatistiken
x_bar = mean(xy[,1]) # Stichprobenmittel der x,...,x_n
y_bar = mean(xy[,2]) # Stichprobenmittel der y_1,...,y_n
s2X = var(xy[,1]) # Stichprobenvarianz der x,...,x_n
s2Y = var(xy[,2]) # Stichprobenvarianz der y_1,...,y_n
c_xy = cov(xy[,1],xy[,2]) # Stichprobenkovarianz

# Stichprobenregressionsgeradenparameter
b_1 = c_xy/s2X # Stichprobenregressionskoeffizient
b_0 = -b_1*x_bar + y_bar # Stichprobenregressionskoeffizient

> beta_0 : 0
> beta_1 : 0.5
> b_0 : 0.136
> b_1 : 0.514
```


Simulation einer Regressionsgerade



Grundlagen

Korrelation und linear-affine Abhängigkeit

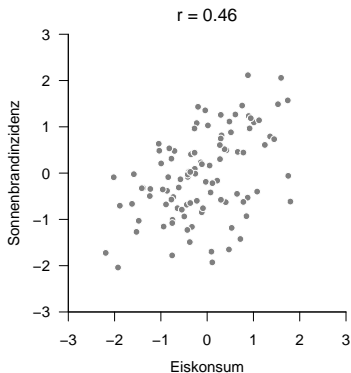
Korrelation und Regression

Korrelation und Bestimmtheitsmaß

Bedingte Korrelation und Partielle Korrelation

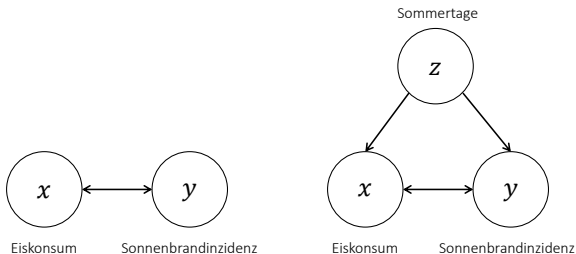
Selbstkontrollfragen

Jährlicher Eiskonsum und jährliche Sonnenbrandinzidenz



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

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 ?

⇒ Bedingte Korrelation und Partielle Korrelation im Falle dreier Zufallsvariablen.

Definition (Bedingte Kovarianz und bedingte Korrelation)

Gegeben seien drei Zufallsvariablen x, y, z einer gemeinsamen Verteilung $\mathbb{P}_{x,y,z}(x, y, z)$. Weiterhin sei $\mathbb{P}_{x,y|z}(x, y)$ die bedingte Verteilung von x und y gegeben z . Dann heißt die Kovarianz von x und y in der Verteilung $\mathbb{P}_{x,y|z}(x, y)$ die *bedingte Kovarianz von x und y gegeben z* und wird mit $\mathbb{C}(x, y|z)$ bezeichnet. Weiterhin seien $\mathbb{P}_{x,y|z}(y)$ und $\mathbb{P}_{x,y|z}(x)$ die marginalen Verteilungen von x und y gegeben z , respektive, und $\mathbb{S}(x|z)$, $\mathbb{S}(y|z)$ die Standardabweichungen von x und y hinsichtlich $\mathbb{P}_{x,y|z}(y)$ und $\mathbb{P}_{x,y|z}(x)$, respektive. Dann heißt die Korrelation von x und y in der Verteilung $\mathbb{P}_{x,y|z}(x, y)$,

$$\rho(x, y|z) := \frac{\mathbb{C}(x, y|z)}{\mathbb{S}(y|z)\mathbb{S}(x|z)} \quad (68)$$

die *bedingte Korrelation von x und y gegeben z*

Bemerkungen

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

Bedingte Korrelation und Partielle Korrelation

Beispiel

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

$$v \sim N(\mu, \Sigma) \quad (69)$$

mit

$$\mu := \begin{pmatrix} \mu_y \\ \mu_x \\ \mu_z \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 & \sigma_{x,z}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 & \sigma_{y,z}^2 \\ \sigma_{z,x}^2 & \sigma_{z,y}^2 & \sigma_z^2 \end{pmatrix} \quad (70)$$

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

$$\Sigma_{x,y} := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 \end{pmatrix}, \Sigma_z := (\sigma_z^2) \text{ und } \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^T := \begin{pmatrix} \sigma_{x,z}^2 \\ \sigma_{y,z}^2 \end{pmatrix}, \quad (71)$$

so dass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix} \quad (72)$$

Mit dem Theorem zu bedingten Normalverteilungen (vgl. (4) Normalverteilungen) ist dann die Kovarianzmatrix des Zufallsvektors (x, y) gegeben durch

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z} \Sigma_z^{-1} \Sigma_{z,(x,y)}. \quad (73)$$

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

$$\Sigma_{x,y|z} = \begin{pmatrix} \mathbb{C}(x, x|z) & \mathbb{C}(x, y|z) \\ \mathbb{C}(y, x|z) & \mathbb{C}(y, y|z) \end{pmatrix}. \quad (74)$$

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

$$\rho(x, y|z) = \frac{\mathbb{C}(x, y|z)}{\sqrt{\mathbb{C}(x, x|z)} \sqrt{\mathbb{C}(y, y|z)}} \quad (75)$$

Für

$$\Sigma := \begin{pmatrix} 1.0 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.5 \\ 0.9 & 0.5 & 1.0 \end{pmatrix} \quad (76)$$

ergibt sich beispielsweise

$$\rho(x, y) = 0.50 \text{ und } \rho(x, y|z) \approx 0.13. \quad (77)$$

Beispiel (fortgeführt)

```
# Bedingte Korrelation bei Normalverteilung
S      = matrix(c( 1,.5,.9,
                  .5, 1,.5,
                  .9,.5, 1), nrow = 3, byrow = TRUE)
# \Sigma

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[2,2]) %*% S[3,1:2]
# \Sigma_{x,y|z}
rho_xy_z = S_xy_z[1,2]/(sqrt(S_xy_z[1,1])*sqrt(S_xy_z[2,2]))
# \rho(x,y|z)

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

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


Definition (Partielle Korrelation)

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

$$\begin{aligned}x &:= \beta_0^{x,z} + \beta_1^{x,z} z \\y &:= \beta_0^{y,z} + \beta_1^{y,z} z\end{aligned}\tag{78}$$

mit Residualvariablen

$$\begin{aligned}e^{x,z} &:= x - \beta_0^{x,z} - \beta_1^{x,z} z \\e^{y,z} &:= y - \beta_0^{y,z} - \beta_1^{y,z} z\end{aligned}\tag{79}$$

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

$$\rho(x, y \setminus z) := \rho(e^{x,z}, e^{y,z}).\tag{80}$$

Bemerkungen

- $e^{x,z}$ ist die Zufallsvariable x , aus der der Einfluss von z "herausgerechnet" wurde.
- $e^{y,z}$ 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 , aus denen jeweils der Einfluss von z "herausgerechnet" wurde

Definition (Partielle Stichprobenkorrelation)

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

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

Schließlich seien für $i = 1, \dots, n$

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

die Residualwerte der jeweiligen Ausgleichsgeraden. Dann heißt die Stichprobenkorrelation der Wertemenge $\{(e_i^{y,z}, e_i^{x,z})\}_{i=1, \dots, n}$ *partielle Stichprobenkorrelation der x_i und y_i mit auspartiellierten 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

$$\rho(x, y|z) = \rho(x, y \setminus z) \quad (81)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Generell sind bedingte und partielle Korrelationen nicht identisch.
- Für Details, siehe zum Beispiel Lawrance (1976) und Baba, Shibata, and Sibuya (2004).

Theorem (Bedingte Korrelation und Korrelationen bei Normalverteilung)

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

$$\rho(x, y|z) = \frac{\rho(x, y) - \rho(x, z)\rho(y, z)}{\sqrt{(1 - \rho(x, z)^2)}\sqrt{(1 - \rho(y, z)^2)}} \quad (82)$$

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

$$r_{x,y|z} = \frac{r_{x,y} - r_{x,z}r_{y,z}}{\sqrt{(1 - r_{x,z}^2)}\sqrt{(1 - r_{y,z}^2)}} \quad (83)$$

- Mit $\rho(x, y|z) = \rho(x, y \setminus z)$ bei Normalverteilung die Formel auch für $\rho(x, y \setminus z)$.

Bedingte Korrelation und Partielle Korrelation

Beweis

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

$$\Sigma := \begin{pmatrix} 1 & \rho(x, y) & \rho(x, z) \\ \rho(y, x) & 1 & \rho(y, z) \\ \rho(z, x) & \rho(z, y) & 1 \end{pmatrix}. \quad (84)$$

Wir definieren nun zunächst

$$\Sigma_{x,y} := \begin{pmatrix} 1 & \rho(x, y) \\ \rho(y, x) & 1 \end{pmatrix}, \Sigma_z := (1) \quad \text{und} \quad \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^T := \begin{pmatrix} \rho(x, z) \\ \rho(y, z) \end{pmatrix}, \quad (85)$$

so dass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix}. \quad (86)$$

Mit dem Theorem zu bedingten Normalverteilungen (vgl. (4) Normalverteilungen) ist dann die Kovarianzmatrix des Zufallsvektors (x, y) gegeben durch

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z} \Sigma_z^{-1} \Sigma_{z,(x,y)}. \quad (87)$$

Beweis (fortgeführt)

Es ergibt sich also

$$\begin{aligned} \begin{pmatrix} \sigma_{x,x|z}^2 & \sigma_{x,y|z}^2 \\ \sigma_{y,x|z}^2 & \sigma_{y,y|z}^2 \end{pmatrix} &= \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x,z) \\ \rho(y,z) \end{pmatrix} (1)^{-1} \begin{pmatrix} \rho(x,z) & \rho(y,z) \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x,z)\rho(x,z) & \rho(x,z)\rho(y,z) \\ \rho(y,z)\rho(x,z) & \rho(y,z)\rho(y,z) \end{pmatrix} \\ &= \begin{pmatrix} 1 - \rho(x,z)^2 & \rho(x,y) - \rho(x,z)\rho(y,z) \\ \rho(y,x) - \rho(y,z)\rho(x,z) & 1 - \rho(y,z)^2 \end{pmatrix}. \end{aligned} \quad (88)$$

Es ergibt sich also

$$\rho(x,y|z) = \frac{\sigma_{x,y|z}^2}{\sqrt{\sigma_{x,x|z}^2} \sqrt{\sigma_{y,y|z}^2}} = \frac{\rho(x,y) - \rho(x,z)\rho(y,z)}{\sqrt{1 - \rho(x,z)^2} \sqrt{1 - \rho(y,z)^2}}. \quad (89)$$

□

Bedingte Korrelation und Partielle Korrelation

Beispiel

```
# Modellformulierung und Datenrealisierung
library(MASS) # Multivariate Normalverteilung
set.seed(1) # reproduzierbare Daten
S = matrix(c( 1,.5,.9, # Kovarianzmatrixparameter \Sigma
            .5, 1,.5,
            .9,.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 Residualstichprobenkorrelation
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 # Residualwerte e^{x,z}
e_yz = xyz[,2] - b_yz1*xyz[,3] - b_yz0 # Residualwerte e^{y,z}
pr_e = cor(e_xz,e_yz) # \rho(x,y|z)

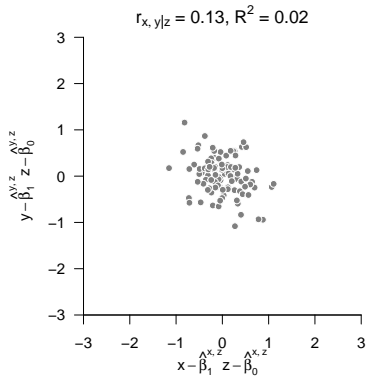
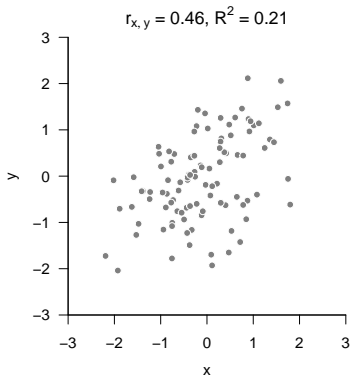
# Partielle Stichprobenkorrelation aus Stichprobenkorrelationen
r = cor(xyz) # Stichprobenkorrelationsmatrix
pr_r_n = r[1,2]-r[1,3]*r[2,3] # \rho(x,y|z) Formel Zähler
pr_r_d = sqrt((1-r[1,3]^2)*(1-r[2,3]^2)) # \rho(x,y|z) Formel Nenner
pr_r = pr_r_n/pr_r_d # \rho(x,y|z)

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

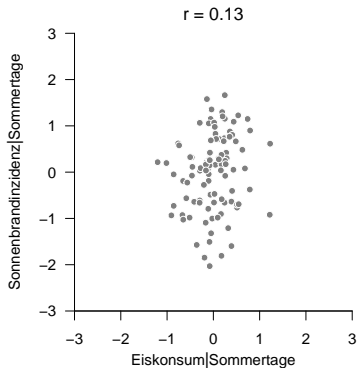
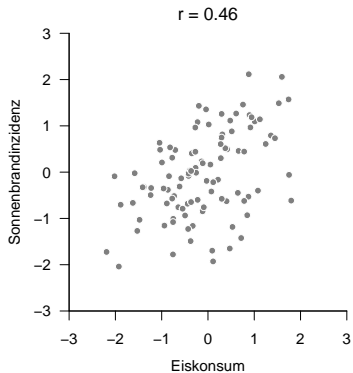
# Ausgabe
cat("r(x,y) : ", r[1,2],
    "\nr(x,y/z) aus Residuenkorrelation : ", pr_e,
    "\nr(x,y/z) aus Korrelationen : ", pr_r,
    "\nr(x,y/z) aus Toolbox : ", pr_t$estimate[1,2])

> 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
```

Bedingte Korrelation und Partielle Korrelation



Bedingte Korrelation und Partielle Korrelation



Grundlagen

Korrelation und Bestimmtheitsmaß

Korrelation und linear-affine Abhängigkeit

Korrelation und Regression

Bedingte Korrelation und Partielle Korrelation

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition der Korrelation zweier Zufallsvariablen wieder.
2. Geben Sie die Definitionen von Stichprobenmittel, -standardabweichung, -kovarianz und -korrelation wieder.
3. Erläutern Sie anhand der Mechanik der Kovariationsterme, wann eine Stichprobenkorrelation einen hohen absoluten Wert annimmt, einen hohen positiven Wert annimmt, einen hohen negativen Wert annimmt und einen niedrigen Wert annimmt.
4. Berechnen Sie die Korrelation von Anzahl der Therapiestunden und Symptomreduktion anhand der Daten in Beispieldatensatz.csv.
5. Geben Sie das Theorem zur Stichprobenkorrelation bei linear-affinen Transformationen wieder.
6. Erläutern Sie das Theorem zur Stichprobenkorrelation bei linear-affinen Transformationen.
7. Geben Sie die Definitionen von erklärten Werten und Residuen einer Ausgleichsgerade wieder.
8. Geben Sie das Theorem zur Quadratsummenzerlegung bei einer Ausgleichsgerade wieder.
9. Erläutern Sie die intuitiven Bedeutungen von SQT, SQE und SQR.
10. Geben Sie die Definition des Bestimmtheitsmaßes R^2 wieder.
11. Geben Sie das Theorem zum Zusammenhang von Stichprobenkorrelation und Bestimmtheitsmaß wieder.
12. Erläutern Sie die Bedeutung von hohen und niedrigen R^2 Werten im Lichte der Ausgleichsgerade.
13. Berechnen Sie in einem R-Skript R^2 für die Daten in der Datei Beispieldatensatz.csv anhand der Definition von R^2 . Überprüfen Sie Ihr Ergebnis anhand des Theorems zum Zusammenhang von Stichprobenkorrelation und Bestimmtheitsmaß.
14. Geben Sie das Theorem zum Zusammenhang von Korrelation und linear-affiner Abhängigkeit wieder.

References

- 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.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(3) Matrizen

Motivation

Matrizen sind die Worte der Sprache des Allgemeinen Linearen Modells

Die Modellformulierung des Allgemeinen Lineare Modells basiert auf Matrizen:

- Experimentelle Designs sind spezielle Matrizen.
- Datensätze sind spezielle Matrizen.
- Modellparameter sind spezielle Matrizen.
- Probabilistische Annahmen sind in Kovarianzmatrizen kodiert.

Modellschätzung und Modellevaluation geschieht mithilfe der Matrizenrechnung.

Ein sicherer Umgang mit Matrizen ist für
das Verständnis des Allgemeinen Linearen Modells unverzichtbar.

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Definition (Matrix)

Eine Matrix ist eine rechteckige Anordnung von Zahlen, die wie folgt bezeichnet wird

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}. \quad (1)$$

Bemerkungen

- Matrizen bestehen aus *Zeilen (rows)* und *Spalten (columns)*.
- Die Matrixeinträge a_{ij} werden mit einem *Zeilenindex* i und einem *Spaltenindex* j indiziert.

- Zum Beispiel gilt für $A := \begin{pmatrix} 2 & 7 & 5 & 2 \\ 8 & 2 & 5 & 6 \\ 6 & 4 & 0 & 9 \\ 9 & 2 & 1 & 2 \end{pmatrix}$, dass $a_{32} = 4$.

Bemerkungen (fortgeführt)

- Die *Größe* oder *Dimension* einer Matrix ergibt sich aus der Anzahl ihrer Zeilen $n \in \mathbb{N}$ und Spalten $m \in \mathbb{N}$.
- Matrizen mit $n = m$ heißen *quadratische Matrizen*.
- In der Folge benötigen wir nur Matrizen mit reellen Einträgen, also $a_{ij} \in \mathbb{R} \forall i = 1, \dots, n, j = 1, \dots, m$.
- Wir nennen die Matrizen mit reellen Einträge *reelle Matrizen*.
- Die Menge der reellen Matrizen mit n Zeilen und m Spalten bezeichnen wir mit $\mathbb{R}^{n \times m}$
- Aus dem Ausdruck $A \in \mathbb{R}^{2 \times 3}$ lesen wir ab, dass A eine reelle Matrix mit zwei Zeilen und drei Spalten ist.
- Wir identifizieren die Menge $\mathbb{R}^{1 \times 1}$ mit der Menge \mathbb{R} .
- Wir identifizieren die Menge $\mathbb{R}^{n \times 1}$ mit der Menge \mathbb{R}^n .
- Reelle Matrizen mit einer Spalte und n Zeilen sind also dasselbe wie n -dimensionale reelle Vektoren.

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Matrixoperationen

Man kann mit Matrizen rechnen.

In der Folge betrachten wir folgende grundlegende Matrixoperationen

- Addition und Subtraktion von Matrizen gleicher Größe (Matrixaddition und Matrixsubtraktion)
- Multiplikation einer Matrix mit einem Skalar (Skalarmultiplikation)
- Vertauschen der Zeilen- und Spaltenanordnung (Matrixtransposition)
- Multiplikation einer Matrix mit einer passenden zweiten Matrix (Matrixmultiplikation)
- "Teilen" durch eine Matrix (Matrixinversion)

Definition (Matrixaddition)

Es seien $A, B \in \mathbb{R}^{n \times m}$. Dann ist die *Addition* von A und B definiert als die Abbildung

$$+ : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto +(A, B) := A + B \quad (2)$$

mit

$$\begin{aligned} A + B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\ &:= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}. \end{aligned} \quad (3)$$

Bemerkungen

- Nur Matrizen identischer Größe können miteinander addiert werden.
- Die Addition zweier gleich großer Matrizen ist elementweise definiert.

Definition (Matrixsubtraktion)

Es seien $A, B \in \mathbb{R}^{n \times m}$. Dann ist die *Subtraktion* von A und B definiert als die Abbildung

$$- : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto -(A, B) := A - B \quad (4)$$

mit

$$\begin{aligned} A - B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\ &:= \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \cdots & a_{1m} - b_{1m} \\ a_{21} - b_{21} & a_{22} - b_{22} & \cdots & a_{2m} - b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} - b_{n1} & a_{n2} - b_{n2} & \cdots & a_{nm} - b_{nm} \end{pmatrix}. \end{aligned} \quad (5)$$

Bemerkungen

- Nur Matrizen identischer Größe können voneinander subtrahiert werden.
- Die Subtraktion zweier gleich großer Matrizen ist elementweise definiert.

Operationen

Beispiel

Es seien $A, B \in \mathbb{R}^{2 \times 3}$ definiert als

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix}. \quad (6)$$

Da A und B gleich groß sind, können wir sie addieren

$$\begin{aligned} C = A + B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2+4 & -3+1 & 0+0 \\ 1-4 & 6+2 & 5+0 \end{pmatrix} \\ &= \begin{pmatrix} 6 & -2 & 0 \\ -3 & 8 & 5 \end{pmatrix} \end{aligned} \quad (7)$$

und voneinander subtrahieren

$$\begin{aligned} D = A - B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} - \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2-4 & -3-1 & 0-0 \\ 1+4 & 6-2 & 5-0 \end{pmatrix} \\ &= \begin{pmatrix} -2 & -4 & 0 \\ 5 & 4 & 5 \end{pmatrix}. \end{aligned} \quad (8)$$

Operationen

Beispiel

```
# Spaltenweise Definition von A (R default)
```

```
A = matrix(c(2,1,-3,6,0,5), nrow = 2)
```

```
print(A)
```

```
>      [,1] [,2] [,3]
```

```
> [1,]    2   -3    0
```

```
> [2,]    1    6    5
```

```
# Zeilenweise Definition von B
```

```
B = matrix(c(4,1,0,-4,2,0), nrow = 2, byrow = TRUE)
```

```
print(B)
```

```
>      [,1] [,2] [,3]
```

```
> [1,]    4    1    0
```

```
> [2,]   -4    2    0
```

Operationen

Beispiel

```
# Addition
```

```
C = A + B
```

```
print(C)
```

```
>      [,1] [,2] [,3]
```

```
> [1,]    6  -2    0
```

```
> [2,]   -3    8    5
```

```
# Subtraktion
```

```
D = A - B
```

```
print(D)
```

```
>      [,1] [,2] [,3]
```

```
> [1,]   -2  -4    0
```

```
> [2,]    5    4    5
```

Definition (Skalarmultiplikation)

Es sei $c \in \mathbb{R}$ ein Skalar und $A \in \mathbb{R}^{n \times m}$. Dann ist die *Skalarmultiplikation* von c und A definiert als die Abbildung

$$\cdot : \mathbb{R} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (c, A) \mapsto \cdot(c, A) := cA \quad (9)$$

mit

$$cA = c \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}. \quad (10)$$

Bemerkungen

- Die Skalarmultiplikation ist elementweise definiert.

Beispiel

Es seien $c := -3$ und $A \in \mathbb{R}^{4 \times 3}$ definiert als

$$A := \begin{pmatrix} 3 & 1 & 1 \\ 5 & 2 & 5 \\ 2 & 7 & 1 \\ 3 & 4 & 2 \end{pmatrix}. \quad (11)$$

Dann ergibt sich

$$B := cA = -3 \begin{pmatrix} 3 & 1 & 1 \\ 5 & 2 & 5 \\ 2 & 7 & 1 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} -3 \cdot 3 & -3 \cdot 1 & -3 \cdot 1 \\ -3 \cdot 5 & -3 \cdot 2 & -3 \cdot 5 \\ -3 \cdot 2 & -3 \cdot 7 & -3 \cdot 1 \\ -3 \cdot 3 & -3 \cdot 4 & -3 \cdot 2 \end{pmatrix} = \begin{pmatrix} -9 & -3 & -3 \\ -15 & -6 & -15 \\ -6 & -21 & -3 \\ -9 & -12 & -6 \end{pmatrix}. \quad (12)$$

Operationen

Beispiel

```
# Definitionen
A = matrix(c(3,1,1,
            5,2,5,
            2,7,1,
            3,4,2),
          nrow = 4,
          byrow = TRUE)

c = -3

# Skalarmultiplikation
B = c*A
print(B)
```

```
>      [,1] [,2] [,3]
> [1,]  -9  -3  -3
> [2,] -15  -6 -15
> [3,]  -6 -21  -3
> [4,]  -9 -12  -6
```

Theorem (Vektorraum $\mathbb{R}^{n \times m}$)

Das Tripel $(\mathbb{R}^{n \times m}, +, \cdot)$ mit der oben definierten Matrixaddition und Skalarmultiplikation ist ein Vektorraum. Insbesondere gelten also für $A, B, C \in \mathbb{R}^{n \times m}$ und $r, s, t \in \mathbb{R}$ folgende Rechenregeln:

- | | |
|--|---|
| (1) Kommutativität der Addition | $A + B = B + A$ |
| (2) Assoziativität der Addition | $(A + B) + C = A + (B + C)$ |
| (3) Existenz eines neutralen Elements der Addition | $\exists 0 \in \mathbb{R}^{n \times m}$ mit $A + 0 = 0 + A = A$. |
| (4) Existenz inverser Elemente der Addition | $\forall A \exists -A$ mit $A + (-A) = 0$. |
| (5) Existenz eines neutralen Elements der Skalarmultiplikation | $\exists 1 \in \mathbb{R}$ mit $1 \cdot A = A$. |
| (6) Assoziativität der Skalarmultiplikation | $r \cdot (s \cdot t) = (r \cdot s) \cdot t$. |
| (7) Distributivität hinsichtlich der Matrixaddition | $r \cdot (A + B) = r \cdot A + r \cdot B$. |
| (8) Distributivität hinsichtlich der Skalaraddition | $(r + s) \cdot A = r \cdot A + s \cdot A$. |

Bemerkungen

- Wir verzichten auf einen Beweis.
- Der Beweis ergibt sich mit dem elementweisen Charakter von $+$, $-$, \cdot und den Rechenregeln in $(\mathbb{R}, +, \cdot)$.
- Das neutrale Element der Addition heißt *Nullmatrix*; wir schreiben $0_{nm} := (0)_{1 \leq i \leq n, 1 \leq j \leq m}$ mit $0 \in \mathbb{R}$.
- Die inversen Elemente der Addition sind durch $-A := (-a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$ gegeben.
- Das neutrale Element der Skalarmultiplikation ist $1 \in \mathbb{R}$.

Definition (Matrixtransposition)

Es sei $A \in \mathbb{R}^{n \times m}$. Dann ist die *Transposition* von A definiert als die Abbildung

$$\cdot^T : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{m \times n}, A \mapsto \cdot^T(A) := A^T \quad (13)$$

mit

$$A^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}^T := \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix} \quad (14)$$

Bemerkungen

- Die Matrixtransposition "vertauscht" Zeilen und Spalten.
- Für $A \in \mathbb{R}^{n \times m}$ gilt immer $A^T \in \mathbb{R}^{m \times n}$.
- Für $A \in \mathbb{R}^{1 \times 1}$ gilt immer $A^T = A$.
- Es gilt $(A^T)^T = A$.
- Es gilt $(a_{ii})_{1 \leq i \leq \min(n,m)} = (a_{ii})_{1 \leq i \leq \min(n,m)}^T$
- Matricelemente auf der Hauptdiagonalen einer Matrix bleiben bei Transposition also unberührt.

Beispiel

Es sei $A \in \mathbb{R}^{2 \times 3}$ definiert durch

$$A := \begin{pmatrix} 2 & 3 & 0 \\ 1 & 6 & 5 \end{pmatrix}, \quad (15)$$

Dann gilt $A^T \in \mathbb{R}^{3 \times 2}$ und speziell

$$A^T := \begin{pmatrix} 2 & 1 \\ 3 & 6 \\ 0 & 5 \end{pmatrix}. \quad (16)$$

Weiterhin gilt offenbar $\min(m, n) = 2$ und folglich

$$(a_{11}) = (a_{11})^T \text{ und } (a_{22}) = (a_{22})^T. \quad (17)$$

Operationen

Beispiel

```
# Definition
A = matrix(c(2,3,0,
            1,6,5),
          nrow = 2,
          byrow = TRUE)
print(A)
```

```
>      [,1] [,2] [,3]
> [1,]    2    3    0
> [2,]    1    6    5
```

```
# Transposition
AT = t(A)
print(AT)
```

```
>      [,1] [,2]
> [1,]    2    1
> [2,]    3    6
> [3,]    0    5
```

Definition (Matrixmultiplikation)

Es seien $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times k}$. Dann ist die *Matrixmultiplikation* von A und B definiert als die Abbildung

$$\cdot : \mathbb{R}^{n \times m} \times \mathbb{R}^{m \times k} \rightarrow \mathbb{R}^{n \times k}, (A, B) \mapsto \cdot(A, B) := AB \quad (18)$$

mit

$$AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{pmatrix} \quad (19)$$
$$:= \begin{pmatrix} \sum_{i=1}^m a_{1i}b_{i1} & \sum_{i=1}^m a_{1i}b_{i2} & \cdots & \sum_{i=1}^m a_{1i}b_{ik} \\ \sum_{i=1}^m a_{2i}b_{i1} & \sum_{i=1}^m a_{2i}b_{i2} & \cdots & \sum_{i=1}^m a_{2i}b_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ni}b_{i1} & \sum_{i=1}^m a_{ni}b_{i2} & \cdots & \sum_{i=1}^m a_{ni}b_{ik} \end{pmatrix}$$

Bemerkungen

- Das Matrixprodukt AB ist nur dann definiert, wenn A genau so viele Spalten hat wie B Zeilen.
- Informell gilt für die beteiligten Matrixgrößen immer $(n \times m)(m \times k) = (n \times k)$.
- In AB ist $(AB)_{ij}$ die Summe der multiplizierten i ten Zeilen von A und j ten Spalten von B .
- Zum Berechnen von $(AB)_{ij}$ für $1 \leq i \leq n, 1 \leq j \leq k$ geht man also wie folgt vor:
 1. Man legt in Gedanken die Transposition der i ten Zeile von A über die j te Spalte von B .
 2. Weil A genau m Spalten hat und B genau m Zeilen hat, gibt es zu jedem Element der Zeile aus A ein korrespondierendes Element in der Spalte von B .
 3. Man multipliziert die korrespondierenden Elemente miteinander.
 4. Die Summe dieser Produkte ist dann der Eintrag mit Index ij in AB .
- Die Multiplikation von Matrizen ist im Allgemeinen nicht kommutativ (also meist $AB \neq BA$).

Beispiel

$A \in \mathbb{R}^{2 \times 3}$ und $B \in \mathbb{R}^{3 \times 2}$ seien definiert als

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix}. \quad (20)$$

Wir wollen $C := AB$ und $D := BA$ berechnen.

Mit $n = 2$, $m = 3$ und $k = 2$ wissen wir schon, dass $C \in \mathbb{R}^{2 \times 2}$ und $D \in \mathbb{R}^{3 \times 3}$, weil

$$(2 \times 3)(3 \times 2) = (2 \times 2) \quad (21)$$

und

$$(3 \times 2)(2 \times 3) = (3 \times 3) \quad (22)$$

Es gilt hier also sicher $AB \neq BA$.

Beispiel (fortgeführt)

Es ergibt sich zum einen

$$\begin{aligned}C &= AB \\&= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix} \\&= \begin{pmatrix} 2 \cdot 4 + (-3) \cdot (-1) + 0 \cdot 1 & 2 \cdot 2 + (-3) \cdot 0 + 0 \cdot 3 \\ 1 \cdot 4 + 6 \cdot (-1) + 5 \cdot 1 & 1 \cdot 2 + 6 \cdot 0 + 5 \cdot 3 \end{pmatrix} & (23) \\&= \begin{pmatrix} 8 + 3 + 0 & 4 + 0 + 0 \\ 4 - 6 + 5 & 2 + 0 + 15 \end{pmatrix} \\&= \begin{pmatrix} 11 & 4 \\ 3 & 17 \end{pmatrix}.\end{aligned}$$

Beispiel (fortgeführt)

```
# Definitionen
A = matrix(c(2,-3,0,
            1, 6,5),
          nrow = 2,
          byrow = TRUE)
B = matrix(c( 4,2,
            -1,0,
            1,3),
          nrow = 3,
          byrow = TRUE)

# Matrixmultiplikation
C = A %*% B
print(C)
```

```
>      [,1] [,2]
> [1,]   11   4
> [2,]    3  17
```

Beispiel (fortgeführt)

Es ergibt sich zum anderen

$$\begin{aligned} D &= BA \\ &= \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 4 \cdot 2 + 2 \cdot 1 & 4 \cdot (-3) + 2 \cdot 6 & 4 \cdot 0 + 2 \cdot 5 \\ (-1) \cdot 2 + 0 \cdot 1 & (-1) \cdot (-3) + 0 \cdot 6 & (-1) \cdot 0 + 0 \cdot 5 \\ 1 \cdot 2 + 3 \cdot 1 & 1 \cdot (-3) + 3 \cdot 6 & 1 \cdot 0 + 3 \cdot 5 \end{pmatrix} \\ &= \begin{pmatrix} 8 + 2 & -12 + 12 & 0 + 5 \\ -2 + 0 & 3 + 0 & 0 + 0 \\ 2 + 3 & -3 + 18 & 0 + 15 \end{pmatrix} \\ &= \begin{pmatrix} 10 & 0 & 10 \\ -2 & 3 & 0 \\ 5 & 15 & 15 \end{pmatrix} \end{aligned} \tag{24}$$

Operationen

Beispiel (fortgeführt)

```
# Definitionen
A = matrix(c(2,-3,0,
            1, 6,5),
          nrow = 2,
          byrow = TRUE)
B = matrix(c( 4,2,
            -1,0,
            1,3),
          nrow = 3,
          byrow = TRUE)
```

```
# Matrixmultiplikation
D = B %*% A
print(D)
```

```
>      [,1] [,2] [,3]
> [1,]  10   0  10
> [2,]  -2   3   0
> [3,]   5  15  15
```

```
# Beispiel für eine undefinierte Matrixmultiplikation
E = t(A) %*% B      # (3 x 2)(3 x 2)
```

```
> Error in t(A) %*% B: nicht passende Argumente
```


Theorem (Matrixmultiplikation und Skalarprodukt)

Es seien $x, y \in \mathbb{R}^n$. Dann gilt

$$\langle x, y \rangle = x^T y. \quad (25)$$

Weiterhin seien für $A \in \mathbb{R}^{n \times m}$ für $i = 1, \dots, n$

$$\bar{a}_i := (a_{ji})_{1 \leq j \leq m} \in \mathbb{R}^m \quad (26)$$

die Spalten von A^T und für $B \in \mathbb{R}^{m \times k}$ für $i = 1, \dots, k$

$$\bar{b}_j := (b_{ij})_{1 \leq i \leq m} \in \mathbb{R}^m \quad (27)$$

die Spalten von B , also

$$A^T = \begin{pmatrix} \bar{a}_1 & \bar{a}_2 & \dots & \bar{a}_n \end{pmatrix} \in \mathbb{R}^{m \times n} \text{ und } B = \begin{pmatrix} \bar{b}_1 & \bar{b}_2 & \dots & \bar{b}_k \end{pmatrix} \in \mathbb{R}^{m \times k}. \quad (28)$$

Dann gilt

$$AB = \left(\langle \bar{a}_i, \bar{b}_j \rangle \right)_{1 \leq i \leq n, 1 \leq j \leq k} \quad (29)$$

Bemerkungen

- Der Eintrag $(AB)_{ij}$ entspricht dem Skalarprodukt von i ter Spalte von A^T und j ter Spalte von B .
- Die erste Aussage folgt mit der Identifikation von $\mathbb{R}^n = \mathbb{R}^{n \times 1}$
- Wir verzichten auf einen ausführlichen Beweis.

Motivation für Begriff der Inversen einer quadratischen Matrix

- Es seien $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$ und $b \in \mathbb{R}^n$, A und b seien als bekannt vorausgesetzt, x sei unbekannt.
- Zum Beispiel sei $A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ und $b := \begin{pmatrix} 5 \\ 11 \end{pmatrix}$
- In diesem Fall gilt $Ax = b \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \Leftrightarrow \begin{array}{l} 1x_1 + 2x_2 = 5 \\ 3x_1 + 4x_2 = 11 \end{array}$
- Wir haben also ein *lineares Gleichungssystem (LGS)* mit zwei Gleichungen und zwei Unbekannten.
- Wir stellen uns vor, dass wissen möchten, für welche(s) x das LGS erfüllt ist.
- Wären $A = a \in \mathbb{R}$, $x \in \mathbb{R}$ und $b \in \mathbb{R}$, also $ax = b$ gegeben so würden mit dem *multiplikativen Inversen* von a multiplizieren, also dem Wert, der mit a multipliziert 1 ergibt und durch $a^{-1} = \frac{1}{a}$ gegeben ist.
- Dann würde nämlich gelten $ax = b \Leftrightarrow a^{-1}ax = a^{-1}b \Leftrightarrow 1 \cdot x = a^{-1}b \Leftrightarrow x = \frac{b}{a}$
- Konkret etwa $2x = 6 \Leftrightarrow 2^{-1}2x = 2^{-1}6 \Leftrightarrow \frac{1}{2}2x = \frac{1}{2}6 \Leftrightarrow x = 3$.
- Analog möchte mit dem *multiplikativen Inversen* A^{-1} von A multiplizieren können, sodass " $A^{-1}A = 1$ ".
- Dann hätte man nämlich $Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b$
- Die Idee des multiplikativen Inversen wird im folgenden als *Inverse einer quadratischen Matrix* formalisiert.

Definition (Einheitsmatrix)

Die Matrix

$$I_n := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \quad (30)$$

mit $a_{ij} = 1$ für $i = j$ und $a_{ij} = 0$ für $i \neq j$ heißt *n-dimensionale Einheitsmatrix*.

- I_n wird in R mit dem Befehl `diag(n)` erzeugt.

Theorem (Neutrales Element der Matrixmultiplikation)

I_n ist das neutrale Element der Matrixmultiplikation, d.h. es gilt für $A \in \mathbb{R}^{n \times m}$, dass

$$I_n A = A \text{ und } A I_m = A. \quad (31)$$

Beweis

Es sei $B = (b_{ij}) = I_n A \in \mathbb{R}^{n \times m}$. Dann gilt für alle $1 \leq i \leq n$ und alle $1 \leq j \leq m$

$$d_{ij} = 0 \cdot a_{1j} + 0 \cdot a_{2j} + \cdots + 0 \cdot a_{i-1,j} + 1 \cdot a_{ij} + \cdots + 0 \cdot a_{i+1,j} + 0 \cdot a_{nj} = a_{ij} \quad (32)$$

und analog für $A I_m$. □

Definition (Invertierbare Matrix und inverse Matrix)

Eine quadratische Matrix $A \in \mathbb{R}^{n \times n}$ heißt *invertierbar*, wenn es eine quadratische Matrix $A^{-1} \in \mathbb{R}^{n \times n}$ gibt, so dass

$$A^{-1}A = AA^{-1} = I_n \quad (33)$$

ist. Die Matrix A^{-1} heißt die *inverse Matrix von A*.

Bemerkungen

- Invertierbarkeit und inverse Matrizen beziehen sich nur auf quadratische Matrizen.
- Inverse Matrizen heißen auch einfach *Inverse*.
- Quadratische Matrizen können, müssen aber nicht invertierbar sein.
- Nicht invertierbare Matrizen nennt man *singuläre* Matrizen
- Für $A = a \in \mathbb{R}^{1 \times 1}$ gilt $A^{-1} = \frac{1}{a}$.
- Die Definition sagt nur aus, was eine inverse Matrix ist, nicht wie man sie berechnet.

Beispiel für eine invertierbare Matrix

Die Matrix $A = \begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix}$ ist invertierbar mit inverser Matrix $A^{-1} = \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix}$, denn

$$\begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix} \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix}, \quad (34)$$

wovon man sich durch Nachrechnen überzeugt.

Beispiel für eine nicht-invertierbare Matrix

Die Matrix $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ ist nicht invertierbar, denn wäre B invertierbar, dann gäbe es $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ mit

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (35)$$

Das würde aber bedeuten, dass $0 = 1$ in \mathbb{R} und das ist ein Widerspruch. Also kann B nicht invertierbar sein.

Berechnen inverser Matrizen

- 2×2 bis etwa 5×5 Matrizen kann man prinzipiell per Hand invertieren.
- Dazu lernt man im BSc Mathematik verschiedene Verfahren.
- Wir verzichten auf eine Einführung in die Matrizeninvertierung per Hand.
- Ein kurzes (30 min) Erklärvideo [findet sich hier](#).
- In der Anwendung werden Matrizen standardmäßig numerisch invertiert.
- Matrixinversion ist ein weites Feld in der numerischen Mathematik.
- Es gibt sehr viele Algorithmen zur Invertierung invertierbarer Matrizen.
- Elegant berechnet man inverse Matrizen in R zum Beispiel mit dem Paket `matlib`.

Berechnen inverser Matrizen

```
# Einmalige Installation des R Pakets matlib  
install.packages("matlib")
```

```
# Laden der matlib Funktionen  
library(matlib)
```

```
# Definition  
A = matrix(c(2,1,  
            3,4),  
          nrow = 2,  
          byrow = TRUE)
```

```
# Berechnen von  $A^{-1}$   
inv(A)
```

```
>      [,1] [,2]  
> [1,]  0.8 -0.2  
> [2,] -0.6  0.4
```

Berechnen inverser Matrizen

```
print(inv(A) %*% A)
```

```
>           [,1] [,2]
> [1,] 1.00e+00  0
> [2,] 2.22e-16  1
```

```
print(A %*% inv(A))
```

```
>           [,1] [,2]
> [1,] 1.00e+00  0
> [2,] 4.44e-16  1
```

```
# Nicht-invertierbare Matrizen sind auch numerisch nicht-invertierbar (singular)
```

```
B = matrix(c(1,0,
             0,0),
           nrow = 2,
           byrow = 2)
```

```
inv(B)
```

```
> Error in Inverse(X, tol = sqrt(.Machine$double.eps), ...): X is numerically singular
```

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Definition (Determinante)

Für $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ mit $n > 1$ sei $A_{ij} \in \mathbb{R}^{n-1 \times n-1}$ die Matrix, die aus A durch Entfernen der i ten Zeile und der j ten Spalte entsteht. Dann heißt die Zahl

$$\det(A) := a_{11} \quad \text{für } n = 1 \quad (36)$$

$$\det(A) := \sum_{j=1}^n a_{1j} (-1)^{1+j} \det(A_{1j}) \quad \text{für } n > 1 \quad (37)$$

die *Determinante von A*.

Bemerkungen

- Für

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad (38)$$

ergeben sich zum Beispiel

$$A_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, A_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, A_{21} = \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}, A_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix} \quad (39)$$

- Determinanten sind nichtlineare Abbildungen der Form $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}, A \mapsto \det(A)$

Theorem (Determinanten von 2×2 und 3×3 Matrizen)

(1) Es sei $A = (a_{ij})_{1 \leq i, j \leq 2} \in \mathbb{R}^{2 \times 2}$. Dann gilt

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}. \quad (40)$$

(2) Es sei $A = (a_{ij})_{1 \leq i, j \leq 3} \in \mathbb{R}^{3 \times 3}$. Dann gilt

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \quad (41)$$

Bemerkungen

- Für 2×2 und 3×3 Matrizen (und nur für diese) gilt die *Sarrusche Merkregel*
"Summe der Produkte auf den Diagonalen minus Summe der Produkte auf den Gegendiagonalen"
- Bei 3×3 Matrizen bezieht sich die Merkregel auf das Schema

$$\left(\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \right) \quad (42)$$

Determinanten

Beweis

Für $A \in \mathbb{R}^{2 \times 2}$ gilt nach Definition

$$\begin{aligned}\det(A) &= \sum_{j=1}^n a_{1j} (-1)^{1+j} \det(A_{1j}) \\ &= a_{11} (-1)^{1+1} \det(A_{11}) + a_{12} (-1)^{1+2} \det(A_{12}) \\ &= a_{11} \det((a_{22})) - a_{12} \det((a_{21})) \\ &= a_{11} a_{22} - a_{12} a_{21}\end{aligned}\tag{43}$$

Für $A \in \mathbb{R}^{3 \times 3}$ gilt nach Definition und mit der Formel für Determinanten von 2×2 Matrizen

$$\begin{aligned}\det(A) &= \sum_{j=1}^n a_{1j} (-1)^{1+j} \det(A_{1j}) \\ &= a_{11} (-1)^{1+1} \det(A_{1j}) + a_{12} (-1)^{1+2} \det(A_{12}) + a_{13} (-1)^{1+3} \det(A_{13}) \\ &= a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13}) \\ &= a_{11} \det\left(\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}\right) - a_{12} \det\left(\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix}\right) + a_{13} \det\left(\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}\right) \\ &= a_{11} (a_{22} a_{33} - a_{23} a_{32}) - a_{12} (a_{21} a_{33} - a_{23} a_{31}) + a_{13} (a_{21} a_{32} - a_{22} a_{31}) \\ &= a_{11} a_{22} a_{33} - a_{11} a_{23} a_{32} - a_{12} a_{21} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} - a_{13} a_{22} a_{31} \\ &= a_{11} a_{22} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} - a_{12} a_{21} a_{33} - a_{11} a_{23} a_{32} - a_{13} a_{22} a_{31}.\end{aligned}\tag{44}$$

Determinanten

Beispiel 1

Es seien

$$A := \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \text{ und } B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (45)$$

Dann ergeben sich

$$\det(A) = 2 \cdot 4 - 1 \cdot 3 = 8 - 3 = 5. \quad (46)$$

und

$$\det(B) = 1 \cdot 0 - 0 \cdot 0 = 0 - 0 = 0. \quad (47)$$

Beispiel 2 Es sei

$$C := \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad (48)$$

Dann ergibt sich

$$\det(C) = 2 \cdot 1 \cdot 3 + 0 \cdot 0 \cdot 0 + 0 \cdot 0 \cdot 0 - 0 \cdot 0 \cdot 3 - 0 \cdot 0 \cdot 0 - 0 \cdot 1 \cdot 0 = 2 \cdot 1 \cdot 3 = 6 \quad (49)$$

Determinanten

```
# Beispiel 1  
A = matrix(c(2,1,  
            3,4),  
          nrow = 2,  
          byrow = TRUE)  
  
det(A) # Determinantenberechnung
```

```
> [1] 5  
B = matrix(c(1,0,  
            0,0),  
          nrow = 2,  
          byrow = TRUE)  
  
det(B) # Determinantenberechnung
```

```
> [1] 0  
# Beispiel 2  
C = matrix(c(2,0,0,  
            0,1,0,  
            0,0,3),  
          nrow = 3,  
          byrow = TRUE)  
  
det(C) # Determinantenberechnung
```

```
> [1] 6
```

Theorem (Rechenregeln für Determinanten)

(Determinantenmultiplikationssatz.) Für $A, B \in \mathbb{R}^{n \times n}$ gilt

$$\det(AB) = \det(A) \det(B). \quad (50)$$

(Transposition.) Für $A \in \mathbb{R}^{n \times n}$ gilt

$$\det(A) = \det(A^T). \quad (51)$$

(Inversion.) Für eine invertierbare Matrix $A \in \mathbb{R}^{n \times n}$ gilt

$$\det(A^{-1}) = \frac{1}{\det(A)} \quad (52)$$

(Dreiecksmatrizen.) Für Matrizen $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ mit $a_{ij} = 0$ für $i > j$ oder $a_{ij} = 0$ für $j > i$ gilt

$$\det(A) = \prod_{i=1}^n a_{ii} \quad (53)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Bei Dreiecksmatrizen sind alle Elemente unterhalb ($i > j$) oder oberhalb ($j > i$) der Diagonalen 0
- Bei I_n sind alle nicht-diagonalen Elemente 0 und alle diagonalen Elemente 1, also folgt $\det(I_n) = 1$.

Theorem (Invertierbarkeit und Determinante)

$A \in \mathbb{R}^{n \times n}$ ist dann und nur dann invertierbar, wenn gilt, dass $\det(A) \neq 0$. Es gilt also

$$A \text{ ist invertierbar} \Leftrightarrow \det(A) \neq 0 \text{ und } A \text{ ist nicht invertierbar} \Leftrightarrow \det(A) = 0. \quad (54)$$

Beweisandeutung

Wir zeigen lediglich, dass aus der Invertierbarkeit von A folgt, dass $\det(A)$ nicht null sein kann. Nehmen wir also an, dass A invertierbar ist. Dann gibt es eine Matrix B mit $AB = I_n$ und mit dem Determinantenmultiplikationssatz folgt

$$\det(AB) = \det(A) \det(B) = \det(I_n) = 1. \quad (55)$$

Also kann $\det(A) = 0$ nicht gelten, denn sonst wäre $0 = 1$.

□

Determinanten

Visuelle Intuition

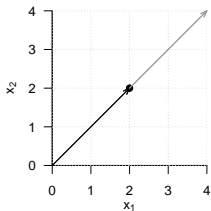
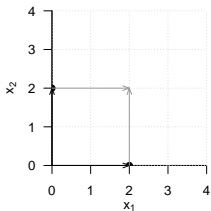
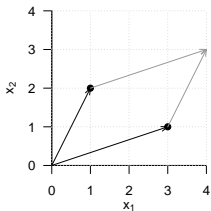
$a_1, \dots, a_n \in \mathbb{R}^n$ seien die Spalten von $A \in \mathbb{R}^{n \times n}$.

$\Rightarrow \det(A)$ entspricht dem signierten Volumen des von $a_1, \dots, a_n \in \mathbb{R}^n$ aufgespannten Parallelotops.

$$A_1 = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$



$$\det(A_1) = 3 \cdot 2 - 1 \cdot 1 = 5$$

$$\det(A_2) = 2 \cdot 2 - 0 \cdot 0 = 4$$

$$\det(A_3) = 2 \cdot 2 - 2 \cdot 2 = 0$$

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Überblick

- Der Rang einer Matrix ist eine Zahl an der bestimmte Eigenschaften der Matrix abgelesen werden können.
- In dieser Hinsicht ist der Rang einer Matrix sehr ähnlich zur Determinante einer Matrix.
- Viele Resultate in der linearen Algebra beruhen auf Annahmen über den Rang einer Matrix.
- Der Rang einer Matrix ist ein tiefgehendes Konzept, das wir hier nur oberflächlich behandeln können.
- Für ausführlichere Einführungen, siehe z.B. Searle (1982), Chapter 6 und Strang (2009), Kapitel Chapter 3.2.
- Wir verwenden hier einen Zugang über das Konzept der linearen Unabhängigkeit von Vektoren.
- Wir erinnern zunächst an dieses Konzept.

Definition (Lineare Unabhängigkeit)

V sei ein Vektorraum. Eine Menge $W := \{w_1, w_2, \dots, w_k\}$ von Vektoren in V heißt *linear unabhängig*, wenn die einzige Repräsentation des Nullelements $0 \in V$ durch eine Linearkombination der $w \in W$ die triviale Repräsentation

$$0 = a_1 w_1 + a_2 w_2 + \dots + a_n w_n \text{ mit } a_1 = a_2 = \dots = a_n = 0 \quad (56)$$

ist. Wenn die Menge W nicht linear unabhängig ist, dann heißt sie *linear abhängig*.

Bemerkungen

- Prinzipiell müsste man für jede Linearkombination der $w \in W$ prüfen, ob sie Null ist.
- Die beiden folgenden Theoreme zeigen, dass es auch einfacher geht.

Theorem (Lineare Abhängigkeit von zwei Vektoren)

V sei ein Vektorraum. Zwei Vektoren $v_1, v_2 \in V$ sind linear abhängig, wenn einer der Vektoren ein skalares Vielfaches des anderen Vektors ist.

Beweis

v_1 sei ein skalares Vielfaches von v_2 , also

$$v_1 = \lambda v_2 \text{ mit } \lambda \neq 0. \quad (57)$$

Dann gilt

$$v_1 - \lambda v_2 = 0. \quad (58)$$

Dies aber entspricht der Linearkombination

$$a_1 v_1 + a_2 v_2 = 0 \quad (59)$$

mit $a_1 = 1 \neq 0$ und $a_2 = -\lambda \neq 0$. Es gibt also eine Linearkombination des Nullelementes, die nicht die triviale Repräsentation ist, und damit sind v_1 und v_2 nicht linear unabhängig.

Theorem (Lineare Abhängigkeit einer Menge von Vektoren)

V sei ein Vektorraum und $w_1, \dots, w_k \in V$ sei eine Menge von Vektoren in V . Wenn einer der Vektoren $w_i, i = 1, \dots, k$ eine Linearkombination der anderen Vektoren ist, dann ist die Menge der Vektoren linear abhängig.

Beweis

Die Vektoren w_1, \dots, w_k sind genau dann linear abhängig, wenn gilt, dass $\sum_{i=1}^n a_i w_i = 0$ mit mindestens einem $a_i \neq 0$. Es sei also zum Beispiel $a_j \neq 0$. Dann gilt

$$0 = \sum_{i=1}^n a_i w_i = \sum_{i=1, i \neq j}^n a_i w_i + a_j w_j \quad (60)$$

Also folgt

$$a_j w_j = - \sum_{i=1, i \neq j}^n a_i w_i \quad (61)$$

und damit

$$w_j = -a_j^{-1} \sum_{i=1, i \neq j}^n a_i w_i = - \sum_{i=1, i \neq j}^n (a_j^{-1} a_i) w_i \quad (62)$$

Also ist w_j eine Linearkombination der $w_i, i = 1, \dots, k$ mit $i \neq j$. □

Definition (Rang einer Matrix)

Es sei $A \in \mathbb{R}^{n \times m}$ und

$$a_1 := \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, a_2 := \begin{pmatrix} a_{12} \\ \vdots \\ a_{n2} \end{pmatrix}, \dots, a_n := \begin{pmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{pmatrix} \in \mathbb{R}^n \quad (63)$$

seien die *Spalten(vektoren)* von A . Dann ist *der Rang* von A , geschrieben als $\text{rg}(A)$ definiert als die maximale Anzahl der linear unabhängigen Spalten(vektoren) von A . Ist die Anzahl der maximal linear unabhängigen Spalten(vektoren) von A gleich m , so sagt man, dass A *vollen Spaltenrang* hat.

Bemerkungen

- Die Spalten einer Matrix werden hier als Vektoren in \mathbb{R}^n verstanden.
- Die Definition macht keine Aussage darüber, wie der Rang einer Matrix zu bestimmen ist.
- Es gibt verschiedene Algorithmen um den Rang einer Matrix zu bestimmen, wir vertiefen dies nicht.

Beispiele

(1) Es sei

$$X := \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad (64)$$

Dann sind die Spaltenvektoren keine skalaren Vielfachen voneinander und damit linear unabhängig. Es gilt also $\text{rg}(X) = 2$.

(2) Es sei

$$X := \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{pmatrix} \quad (65)$$

Dann sind die Spaltenvektoren skalare Vielfache voneinander. Die maximale Möglichkeit aus den Spaltenvektoren linear unabhängige Vektoren auszuwählen ist also 1. Die maximale Anzahl an linear unabhängigen Vektoren der Matrix ist also 1 und es gilt $\text{rg}(X) = 1$.

Bemerkungen

- In Beispiel (2) gerät die Definition des Rangs einer Matrix wie hier gegeben an ihre Grenze.
- Alternative Definitionen, z.B. über die Dimension des Spaltenraumes sind eindeutiger, aber tiefergehend.
- Die einzige Matrix mit Rang 0 ist die Nullmatrix.

Range

Beispiele

```
# Bestimmung des Matrixrangs in R über QR Zerlegung (https://de.wikipedia.org/wiki/QR-Zerlegung)
```

```
# Beispiel (1)
```

```
X = matrix(c(1,0,                                # Matrixdefinition
            0,1,
            0,0),
          nrow = 3,
          byrow = TRUE)
```

```
rg = qr(X)$rank                                # Rangevaluation
```

```
print(rg)                                       # Ausgabe
```

```
> [1] 2
```

```
# Beispiel (2)
```

```
X = matrix(c(1,2,                                # Matrixdefinition
            1,2,
            0,0),
          nrow = 3,
          byrow = TRUE)
```

```
rg = qr(X)$rank                                # Rangevaluation
```

```
print(rg)                                       # Ausgabe
```

```
> [1] 1
```

Bemerkungen

- Wir werden einige Aussagen zum Rang einer Matrix für die ALM Theorie benötigen.
- Wir geben diese Aussagen hier nur an, beweisen sie aber nicht.
- Für Beweise verweisen wir auf die einschlägige Literatur

Theoreme zum Rang einer Matrix

- (1) Für $A \in \mathbb{R}^{n \times n}$ gilt: $\text{rg}(A) = n \Leftrightarrow A$ ist invertierbar.
- (2) Für $A \in \mathbb{R}^{n \times n}$ gilt: $\text{rg}(A) < n \Leftrightarrow A$ ist nicht invertierbar.

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Definition (Nullmatrizen, Einheitsmatrizen, Einheitsvektoren, Einsvektoren)

- Wir bezeichnen *Nullmatrizen* mit

$$0_{nm} := (0)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n \quad (66)$$

- Wir bezeichnen die *Einheitsmatrix* mit

$$I_n := (i_{jk})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ f\"ur } j = k \text{ und } i_{jk} = 0 \text{ f\"ur } j \neq k \quad (67)$$

- Wir bezeichnen die *Einheitsvektoren* e_i , $i = 1, \dots, n$ mit

$$e_i := (e_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{ij} = 1 \text{ f\"ur } i = j \text{ und } e_{ij} = 0 \text{ f\"ur } i \neq j \quad (68)$$

- Wir bezeichnen den *Einsvektor* mit

$$1_n := (1)_{1 \leq i \leq n} \in \mathbb{R}^n \quad (69)$$

Bemerkungen

- 0_{nm} und 0_n bestehen nur aus Nullen.
- I_n besteht nur aus Nullen und Diagonalelementen gleich Eins.
- e_i , $i=1, \dots, n$ besteht nur aus Nullen und einer Eins in der i ten Komponente.
- 1_n besteht nur aus Einsen.

Definition (Symmetrische, diagonale, und orthogonale Matrizen)

- Eine Matrix $S \in \mathbb{R}^{n \times n}$ heißt *symmetrisch*, wenn gilt dass $S^T = S$.
- Eine Matrix $D \in \mathbb{R}^{n \times n}$ heißt *Diagonalmatrix*, wenn $d_{ij} = 0$ für $1 \leq i, j \leq n, i \neq j$.
- Eine Matrix $Q \in \mathbb{R}^{n \times n}$ heißt *orthogonal*, wenn ihre Spaltenvektoren wechselseitig *orthonormal* sind.

Bemerkungen

- Eine Diagonalmatrix D mit Diagonalelementen d_1, \dots, d_n schreibt man auch als $D = \text{diag}(d_1, \dots, d_n)$.
- Symmetrische, diagonale, und orthogonale Matrizen haben viele "gute" Eigenschaften.
- Zum Beispiel überzeugt man sich einfach davon, dass Multiplikation einer Matrix A von links mit einer Diagonalmatrix D der Multiplikation der Zeilen der Matrix A mit den entsprechenden Diagonaleinträgen von D entspricht. Die entsprechende Multiplikation von rechts entspricht der Multiplikation der Spalten von A mit entsprechenden Diagonaleinträgen von D .
- Eine weitere im Folgenden wichtige Eigenschaft von Diagonalmatrizen ist
 - $D := \text{diag}(d_1, \dots, d_n)$ ist eine Diagonalmatrix $\Rightarrow \det(D) = \prod_{i=1}^n d_i$.

Definition (Positiv-definite Matrix)

Eine quadratische Matrix $C \in \mathbb{R}^{n \times n}$ heißt positiv-definit (p.d.), wenn

- C eine symmetrische Matrix ist und
- für alle $x \in \mathbb{R}^n, x \neq 0_n$ gilt, dass $x^T C x > 0$ ist.

Bemerkungen

- Im ALM Kontext sind p.d. Matrizen für die Definition der multivariaten Normalverteilungen grundlegend.
 - Wir werden einige Aussagen zu positiv-definiten Matrizen für die ALM Theorie benötigen.
 - Wir halten dieses Aussagen hier ohne Beweis fest und verweisen für Beweise auf die einschlägige Literatur
- (1) Jede positiv-definite Matrix ist invertierbar.
 - (2) Die Inverse einer positiv-definiten Matrix ist ebenfalls positiv-definit.

Definition

Operationen

Determinanten

Rang

Spezielle Matrizen

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie Definition einer Matrix wieder.
2. Nennen Sie sechs Matrixoperationen.
3. Geben Sie Definitionen der Matrixaddition und -subtraktion wieder.
4. Geben Sie die Definition der Skalarmultiplikation für Matrizen wieder.
5. Geben Sie die Definition der Matrixtransposition wieder.

6. Es seien

$$A := \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, B := \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix}, \text{ und } c := 2 \quad (70)$$

Berechnen Sie

$$D := c(A - B^T) \text{ und } E := (cA)^T + B. \quad (71)$$

per Hand und überprüfen Sie Ihre Rechnung mit R.

7. Geben Sie die Definition der Matrixmultiplikation wieder.
8. Es seien $A \in \mathbb{R}^{3 \times 2}$, $B \in \mathbb{R}^{2 \times 4}$ und $C \in \mathbb{R}^{3 \times 4}$. Prüfen Sie, ob folgende Matrixprodukte definiert sind, und wenn ja, geben Sie die Größe der resultierenden Matrix an

$$ABC, \quad ABC^T, \quad , A^T C B^T \quad , BAC \quad (72)$$

9. Es seien

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 2 & 0 \end{pmatrix} \quad B := \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 0 & 0 \end{pmatrix} \quad \text{und} \quad C := \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \quad (73)$$

Berechnen Sie die Matrixprodukte

$$AB, \quad B^T A^T, \quad (B^T A^T)^T, \quad AC \quad (74)$$

per Hand und überprüfen Sie Ihre Rechnung mit R.

10. Invertieren Sie die Matrizen A und B aus der vorherigen Aufgabe mithilfe von `matlib::inv` und überprüfen Sie die Inverseeigenschaft der inversen Matrizen mithilfe von R.
11. Geben Sie die Formel für die Determinante von $A := (A_{ij})_{1 \leq i, j \leq 2} \in \mathbb{R}^2$ wieder.
12. Geben Sie die Formel für die Determinante von $A := (A_{ij})_{1 \leq i, j \leq 3} \in \mathbb{R}^3$ wieder.
13. Berechnen Sie die Determinanten von

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad B := \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{und} \quad C := \text{diag}(1, 2, 3) \quad (75)$$

per Hand und überprüfen Sie Ihre Rechnung mit R.

14. Geben Sie die Definition einer symmetrischen Matrix wieder.
15. Geben Sie die Definition einer Diagonalmatrix wieder.
16. Geben Sie die Definition einer orthogonalen Matrix wieder.
17. Geben Sie die Definition einer positiv-definiten Matrix wieder.
18. Geben Sie die Definition des Rangs einer Matrix wieder
19. Wann sagt man, dass eine Matrix vollen Spaltenrang hat?
20. Bestimmen Sie den Rang folgender Matrizen durch überlegen und mithilfe von R

$$X_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, X_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, X_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (76)$$

References

Searle, Shayle. 1982. *Matrix Algebra Useful for Statistics*. Wiley-Interscience.

Strang, Gilbert. 2009. *Introduction to Linear Algebra*.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(4) Normalverteilungen

Die multivariate Normalverteilung ist zentraler Bestandteil des ALMs.

Die multivariate Normalverteilung ist die multivariate Generalisierung der univariaten Normalverteilung.

Die Motivation von Normalverteilungsannahmen liegt immer im Zentralen Grenzwertsatz:

- Probabilistische Terme repräsentieren die Summation sehr vieler Prozesse, die durch die deterministischen Bestandteile eines Modell, also eine mechanistische wissenschaftliche Theorie, nicht erklärt werden.
- Nach dem Zentralen Grenzwertsatz ist die Summe dieser Prozesse normalverteilt.

Darüberhinaus hat die Normalverteilung günstige mathematische Eigenschaften, die auch in der Quantifikation subjektiver Unsicherheit genutzt werden können. Zur Erarbeitung der Inhalte dieser Einheit bietet sich eine Revision von (5) Multivariate Verteilungen aus Wahrscheinlichkeitstheorie und Frequentistische Inferenz an

Definition und Eigenschaften

Transformation

Unabhängigkeit

Gemeinsame Verteilungen

Selbstkontrollfragen

Definition und Eigenschaften

Transformationen

Unabhängigkeit

Gemeinsame Verteilungen

Selbstkontrollfragen

Definition (Multivariate Normalverteilung)

y sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum \mathbb{R}^n und WDF

$$p : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, v \mapsto p(v) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v - \mu)^T \Sigma^{-1}(v - \mu)\right). \quad (1)$$

Dann sagen wir, dass y einer *multivariaten (oder n -dimensionalen) Normalverteilung* mit Erwartungswertparameter $\mu \in \mathbb{R}^n$ und positiv-definitem Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{n \times n}$ unterliegt und nennen y einen (*multivariat*) *normalverteilten Zufallsvektor*. Wir kürzen dies mit $y \sim N(\mu, \Sigma)$ ab. Die WDF eines multivariat normalverteilten Zufallsvektors bezeichnen wir mit

$$N(v; \mu, \Sigma) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v - \mu)^T \Sigma^{-1}(v - \mu)\right). \quad (2)$$

Bemerkungen

- Man beachte, dass wir für Zufallsvektoren von nun an auch kleine Buchstaben zulassen.
- v ist das kleine griechische Ypsilon.
- Die Dimensionalität von $(v - \mu)^T \Sigma^{-1}(v - \mu)$ ergibt sich zu $(1 \times n)(n \times n)(n \times 1) = 1$
- Es gilt weiterhin $(v - \mu)^T \Sigma^{-1}(v - \mu) > 0$, weil auch Σ^{-1} positiv-definit ist.
- Der Parameter $\mu \in \mathbb{R}^n$ entspricht dem Wert höchster Wahrscheinlichkeitsdichte
- Der Term $(2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}$ ist die Normalisierungskonstante für den Exponentialfunktionsterm.

Definition (Erwartungswert und Kovarianzmatrix von Zufallsvektoren)

y sei ein n -dimensionaler Zufallsvektor. Dann ist der *Erwartungswert* von y definiert als der n -dimensionale Vektor

$$\mathbb{E}(y) := (\mathbb{E}(y_1), \dots, \mathbb{E}(y_n))^T \quad (3)$$

und die *Kovarianzmatrix* von y ist definiert als die $n \times n$ Matrix

$$C(y) := \mathbb{E} \left((y - \mathbb{E}(y))(y - \mathbb{E}(y))^T \right) \quad (4)$$

Bemerkungen

- Der Erwartungswert von y ist der Vektor der Erwartungswerte $\mathbb{E}(y_1), \dots, \mathbb{E}(y_n)$.
- Die Kovarianzmatrix ist formal analog zur Kovarianz zweier Zufallsvariablen definiert.

Theorem (Kovarianzmatrix eines Zufallsvektors)

y sei ein n -dimensionaler Zufallsvektor und $\mathbb{C}(y)$ sei seine Kovarianzmatrix. Dann gilt

$$\mathbb{C}(y) = \left(\mathbb{C}(y_i, y_j) \right)_{1 \leq i, j \leq n} = \begin{pmatrix} \mathbb{C}(y_1, y_1) & \mathbb{C}(y_1, y_2) & \cdots & \mathbb{C}(y_1, y_n) \\ \mathbb{C}(y_2, y_1) & \mathbb{C}(y_2, y_2) & \cdots & \mathbb{C}(y_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}(y_n, y_1) & \mathbb{C}(y_n, y_2) & \cdots & \mathbb{C}(y_n, y_n) \end{pmatrix}. \quad (5)$$

Bemerkung

- Die Kovarianzmatrix $\mathbb{C}(y)$ ist also die Matrix der Kovarianzen der Komponenten von y .

Definition und Eigenschaften

Beweis

Es gilt

$$\begin{aligned} \mathbb{C}(y) &:= \mathbb{E} \left((y - \mathbb{E}(y))(y - \mathbb{E}(y))^T \right) \\ &= \mathbb{E} \left(\left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \mathbb{E}(y_1) \\ \vdots \\ \mathbb{E}(y_n) \end{pmatrix} \right) \left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} \mathbb{E}(y_1) \\ \vdots \\ \mathbb{E}(y_n) \end{pmatrix} \right)^T \right) \\ &= \mathbb{E} \left(\begin{pmatrix} y_1 - \mathbb{E}(y_1) \\ \vdots \\ y_n - \mathbb{E}(y_n) \end{pmatrix} \begin{pmatrix} y_1 - \mathbb{E}(y_1) \\ \vdots \\ y_n - \mathbb{E}(y_n) \end{pmatrix}^T \right) \\ &= \mathbb{E} \left(\begin{pmatrix} y_1 - \mathbb{E}(y_1) \\ \vdots \\ y_n - \mathbb{E}(y_n) \end{pmatrix} (y_1 - \mathbb{E}(y_1) \quad \dots \quad y_n - \mathbb{E}(y_n)) \right) \\ &= \mathbb{E} \begin{pmatrix} (y_1 - \mathbb{E}(y_1))(y_1 - \mathbb{E}(y_1)) & \dots & (y_1 - \mathbb{E}(y_1))(y_n - \mathbb{E}(y_n)) \\ \vdots & \ddots & \vdots \\ (y_n - \mathbb{E}(y_n))(y_1 - \mathbb{E}(y_1)) & \dots & (y_n - \mathbb{E}(y_n))(y_n - \mathbb{E}(y_n)) \end{pmatrix} \\ &= \left(\mathbb{E} \left((y_i - \mathbb{E}(y_i))(y_j - \mathbb{E}(y_j)) \right) \right)_{1 \leq i, j \leq n} \\ &= \left(\mathbb{C}(y_i, y_j) \right)_{1 \leq i, j \leq n} \end{aligned}$$

Theorem (Erwartungswert eines normalverteilten Zufallsvektors)

$y \sim N(\mu, \Sigma)$ sei ein multivariat normalverteilter Zufallsvektor mit Erwartungswertparameter $\mu \in \mathbb{R}^n$ und Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{n \times n}$ p.d.. Dann gilt

$$\mathbb{E}(y) = \mu. \quad (6)$$

Bemerkungen

- Der Erwartungswert eines normalverteilten Zufallsvektors entspricht seinem Erwartungswertparameter.
- Wir verzichten auf einen Beweis.

Theorem (Kovarianzmatrix eines normalverteilten Zufallsvektors)

$y \sim N(\mu, \Sigma)$ sei ein multivariat normalverteilter Zufallsvektor mit Erwartungswertparameter $\mu \in \mathbb{R}^n$ und Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{n \times n}$ p.d.. Dann gilt

$$\mathbb{C}(y) = \Sigma. \quad (7)$$

Bemerkungen

- Die Kovarianzmatrix eines normalverteilten Zufallsvektors entspricht seinem Kovarianzmatrixparameter.
- Wir verzichten auf einen Beweis.

Theorem (Bivariate Normalverteilung)

Ein 2-dimensionaler normalverteilter Zufallsvektor $y = (y_1, y_2)^T$ mit Ergebnisraum \mathbb{R}^2 , Erwartungswert und Kovarianzmatrixparameter

$$\mu := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2 \text{ und } \Sigma := \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix} \in \mathbb{R}^{2 \times 2} \text{ p.d.}, \quad (8)$$

respektive, hat bei Definition von

$$\rho := \frac{\sigma_{12}^2}{\sigma_{11}^2 \sigma_{22}^2} \quad (9)$$

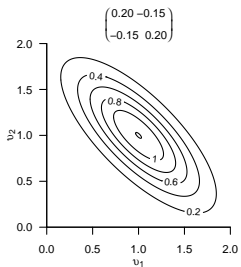
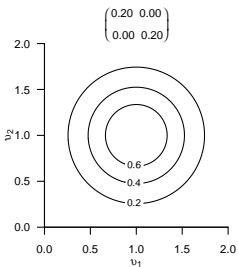
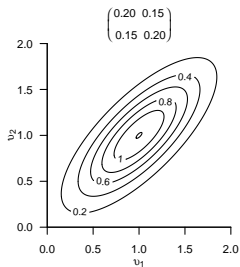
die WDF

$$p : \mathbb{R}^2 \rightarrow \mathbb{R}_{>0}, v \mapsto p(v) = N(v; \mu, \Sigma) \quad (10)$$

mit

$$N(v; \mu, \Sigma) = \frac{1}{2\pi(1 - \rho^2)^{1/2} \sigma_{11} \sigma_{22}} \times \exp \left(-\frac{1}{2(1 - \rho^2)} \left(\left(\frac{v_1 - \mu_1}{\sigma_{11}} \right)^2 - 2\rho \left(\frac{v_1 - \mu_1}{\sigma_{11}} \right) \left(\frac{v_2 - \mu_2}{\sigma_{22}} \right) + \left(\frac{v_2 - \mu_2}{\sigma_{22}} \right)^2 \right) \right)$$

Bivariate Normalverteilungsdichtefunktionen



Definition und Eigenschaften

Beweis

Wir halten zunächst fest, dass aufgrund der positiven Definitheit und damit der Symmetrie von Σ gilt, dass

$$\sigma_{12}^2 = \sigma_{21}^2. \quad (11)$$

Die Inverse von Σ ist gegeben durch

$$\Sigma^{-1} = \frac{1}{\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^4} \begin{pmatrix} \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}^2 \end{pmatrix}, \quad (12)$$

weil

$$\begin{aligned} \Sigma^{-1} \Sigma &= \frac{1}{\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^4} \begin{pmatrix} \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}^2 \end{pmatrix} \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix} \\ &= \frac{1}{\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^4} \begin{pmatrix} \sigma_{22}^2 \sigma_{11}^2 - \sigma_{12}^2 \sigma_{21}^2 & \sigma_{22}^2 \sigma_{12}^2 - \sigma_{12}^2 \sigma_{22}^2 \\ -\sigma_{12}^2 \sigma_{11}^2 + \sigma_{11}^2 \sigma_{21}^2 & -\sigma_{12}^2 \sigma_{12}^2 + \sigma_{11}^2 \sigma_{22}^2 \end{pmatrix} \\ &= \frac{1}{\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^4} \begin{pmatrix} \sigma_{11}^2 \sigma_{11}^2 - \sigma_{12}^4 & 0 \\ 0 & \sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= I_2 \end{aligned} \quad (13)$$

Definition und Eigenschaften

Beweis (fortgeführt)

Mit

$$\rho = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} \Rightarrow \rho = \frac{\sigma_{21}^2}{\sigma_{11}\sigma_{22}} \Rightarrow \sigma_{12}^2 = \sigma_{21}^2 = \rho\sigma_{11}\sigma_{22} \quad (14)$$

lässt sich Σ^{-1} weiterhin schreiben als

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\sigma_{11}^2\sigma_{22}^2 - (\rho\sigma_{11}\sigma_{22})^2} \begin{pmatrix} \sigma_{22}^2 & -\rho\sigma_{11}\sigma_{22} \\ -\rho\sigma_{11}\sigma_{22} & \sigma_{11}^2 \end{pmatrix} \\ &= \frac{1}{\sigma_{11}^2\sigma_{22}^2(1-\rho^2)} \begin{pmatrix} \sigma_{22}^2 & -\rho\sigma_{11}\sigma_{22} \\ -\rho\sigma_{11}\sigma_{22} & \sigma_{11}^2 \end{pmatrix} \end{aligned} \quad (15)$$

Schließlich ergibt sich

$$|\Sigma|^{-\frac{1}{2}} = (|\Sigma|^{-1})^{\frac{1}{2}} = \left| \Sigma^{-1} \right|^{-\frac{1}{2}} = \left(\sigma_{22}^2\sigma_{11}^2 - \rho_{12}^2\sigma_{11}^2\sigma_{22}^2 \right)^{-\frac{1}{2}} = \left(\sigma_{22}\sigma_{11}\sqrt{(1-\rho^2)} \right)^{-1}. \quad (16)$$

Definition und Eigenschaften

Beweis (fortgeführt)

Einsetzen ergibt dann

$$\begin{aligned} N(v; \mu, \Sigma) &= (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v - \mu)^T \Sigma^{-1}(v - \mu)\right) \\ &= (2\pi)^{-\frac{2}{2}} \left(\sigma_{22}\sigma_{11}\sqrt{(1-\rho^2)}\right)^{-1} \\ &\times \exp\left(-\frac{1}{2} \frac{1}{\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^4} \begin{pmatrix} v_1 - \mu_1 & v_2 - \mu_2 \end{pmatrix} \begin{pmatrix} \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}^2 \end{pmatrix} \begin{pmatrix} v_1 - \mu_1 \\ v_2 - \mu_2 \end{pmatrix}\right) \\ &= \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_{11}\sigma_{22}} \\ &\times \exp\left(-\frac{1}{2(1-\rho^2)} \left(\left(\frac{v_1 - \mu_1}{\sigma_{11}}\right)^2 - 2\rho\left(\frac{v_1 - \mu_1}{\sigma_{11}}\right)\left(\frac{v_2 - \mu_2}{\sigma_{22}}\right) + \left(\frac{v_2 - \mu_2}{\sigma_{22}}\right)^2\right)\right), \end{aligned}$$

wobei sich die letzte Gleichung mit der Definition der Matrixmultiplikation ergibt.

Realisierung bivariater normalverteilter Zufallsvektoren

```
# R Paket für multivariate Normalverteilungen  
library(mvtnorm)
```

```
# Parameterdefinition
```

```
mu      = c(1,1)                # \mu in \mathbb{R}^2  
Sigma   = matrix(c(0.2, 0.15, 0.15, 0.2), 2) # \Sigma in \mathbb{R}^{2 \times 2}
```

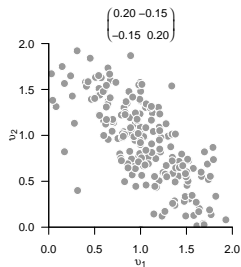
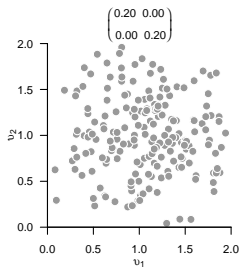
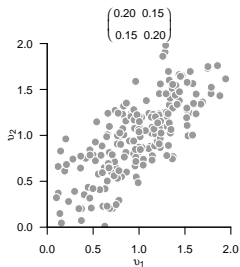
```
# Zufallsvektorrealisierungen
```

```
rmvnorm(n = 10, mu, Sigma)
```

```
>      [,1] [,2]  
> [1,] 0.47631 0.846  
> [2,] 0.00553 0.549  
> [3,] 1.46466 1.582  
> [4,] 0.21191 0.564  
> [5,] 0.84844 0.840  
> [6,] 0.89004 1.154  
> [7,] 1.54162 0.715  
> [8,] 0.86607 0.335  
> [9,] 0.77755 0.883  
> [10,] 1.05314 0.727
```

Definition und Eigenschaften

Realisierung bivariater normalverteilter Zufallsvektoren



Definition

Transformationen

Unabhängigkeit

Gemeinsame Verteilungen

Selbstkontrollfragen

Theorem (Nichtsinguläre lineare Transformation)

$x \sim N(\mu, \Sigma)$ sei ein normalverteilter n -dimensionaler Zufallsvektor und es sei $y := Ax$ mit einer invertierbaren Matrix $A \in \mathbb{R}^{n \times n}$. Dann gilt

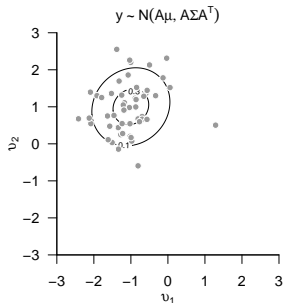
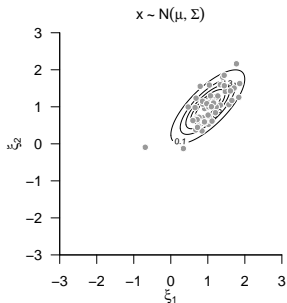
$$y \sim N\left(A\mu, A\Sigma A^T\right) \quad (17)$$

Bemerkung

- Wir verzichten auf einen Beweis.
- Die invertierbare lineare Transformation eines multivariaten normalverteilten Zufallsvektors ergibt wieder einen multivariaten normalverteilten Zufallsvektor. Die Parameter des resultierenden normalverteilten Zufallsvektors ergeben sich dabei aus den Parametern des ursprünglichen Zufallsvektors und der Transformationsmatrix.

Nichtsinguläre lineare Transformation

$$n := 2, \mu := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix}, A := \begin{pmatrix} -2 & -1 \\ 1 & 2 \end{pmatrix}$$



Theorem (Linear-affine Transformation)

$x \sim N(\mu, \Sigma)$ sei ein normalverteilter n -dimensionaler Zufallsvektor und es sei $y := Ax + b$ mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Dann gilt

$$y \sim N(A\mu + b, A\Sigma A^T) \quad (18)$$

Bemerkung

- Für einen Beweis siehe Anderson (2003), Section 2.4.
- Die linear-affine Transformation eines multivariate normalverteilten Zufallsvektors ergibt wieder ein normalverteilten Zufallsvektor. Die Parameter des resultierenden normalverteilten Zufallsvektors ergeben sich dabei aus den Parametern des ursprünglichen Zufallsvektors und der Transformationsparameter.

Definition

Transformationen

Unabhängigkeit

Gemeinsame Verteilungen

Selbstkontrollfragen

Theorem (Unabhängige normalverteilte Zufallsvariablen)

Für $i = 1, \dots, n$ seien $N(v; \mu_i, \sigma^2)$ die WDFen von n unabhängigen univariaten normalverteilten Zufallsvariablen y_1, \dots, y_n mit $\mu_1, \dots, \mu_n \in \mathbb{R}$ und $\sigma^2 > 0$. Weiterhin sei $N(v; \mu, \sigma^2 I_n)$ die WDF eines n -variaten Zufallsvektors y mit Erwartungwertparameter $\mu := (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$. Dann gilt

$$p_y(v) = p_{y_1, \dots, y_n}(v_1, \dots, v_n) \quad (19)$$

und insbesondere

$$N(v; \mu, \sigma^2 I_n) = \prod_{i=1}^n N(v_i; \mu_i, \sigma^2). \quad (20)$$

Bemerkungen

- Das Theorem ist für die Theories des Allgemeinen Linearen Modells zentral.
- Einen Kovarianzmatrixparameter der Form $\sigma^2 I_n$ nennt man auch *sphärisch*.
- Sphärische Kovarianzmatrixparameter von n -variaten Normalverteilungen entsprechen n unabhängigen univariaten Normalverteilungen und umgekehrt. Eine Realisierung einer n -variaten Normalverteilungen entspricht Realisierungen von n unabhängigen univariaten Normalverteilungen.

Unabhängigkeit

Beweis

Wir zeigen die Identität der multivariaten WDF $N(v; \mu, \sigma^2 I_n)$ mit dem Produkt von n univariaten WDFen $N(v_i; \mu_i, \sigma^2 I_n)$, wobei μ_i der i te Eintrag von $\mu \in \mathbb{R}^n$. Es ergibt sich

$$\begin{aligned} N(v; \mu, \sigma^2 I_n) &= (2\pi)^{-\frac{n}{2}} \left| \sigma^2 I_n \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v - \mu)^T (\sigma^2 I_n)^{-1} (v - \mu)\right) \\ &= \left(\prod_{i=1}^n 2\pi^{-\frac{1}{2}} \right) (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (v - \mu)^T (v - \mu)\right) \\ &= \left(\prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (v_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2} (v_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (v_i - \mu_i)^2\right) \\ &= N(v_i; \mu_i, \sigma^2). \end{aligned} \tag{21}$$

Definition

Transformationen

Unabhängigkeit

Gemeinsame Verteilungen

Selbstkontrollfragen

Multivariate Normalverteilungen haben die Eigenschaft, dass auch alle anderen assoziierten Verteilung Normalverteilungen sind und deren Erwartungswert- und Kovarianzmatrixparameter aus den Parametern der jeweils komplementären Verteilung errechnet werden können.

Insbesondere gilt:

- Die uni- und multivariaten Marginalverteilungen multivariater Normalverteilungen sind Normalverteilungen.
- Wie alle multivariaten Verteilungen lassen sich multivariate Normalverteilungen multiplikativ in eine marginale und eine bedingte Verteilung zerlegen. Insbesondere sind bei multivariaten Normalverteilungen diese Verteilungen auch Normalverteilungen, deren Parameter aus den Parametern der gemeinsame Verteilung errechnet werden können und umgekehrt.

Die Resultate dieses Abschnitts sind insbesondere für generalisierte ALMs im Sinne von Linear Mixed Models bzw. Hierarchischen Normalverteilungsmodellen, die Varianzkomponentenschätzung, und die Bayesianische Inferenz bei Normalverteilungsannahmen, aber auch die partielle Korrelation zentral.

Theorem (Marginale Normalverteilungen)

Es sei $n := k + l$ und $x = (x_1, \dots, x_n)$ sei ein n -dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter

$$\mu = \begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix} \in \mathbb{R}^n, \quad (22)$$

mit $\mu_y \in \mathbb{R}^k$ and $\mu_z \in \mathbb{R}^l$ und Kovarianzmatrixparameter

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad (23)$$

mit $\Sigma_{yy} \in \mathbb{R}^{k \times k}$, $\Sigma_{yz} \in \mathbb{R}^{k \times l}$, $\Sigma_{zy} \in \mathbb{R}^{l \times k}$, und $\Sigma_{zz} \in \mathbb{R}^{l \times l}$. Dann sind $y := (x_1, \dots, x_k)$ und $z := (x_{k+1}, \dots, x_n)$ k - und l -dimensionale normalverteilte Zufallsvektoren, respektive, und es gilt

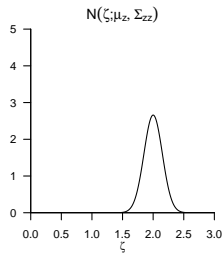
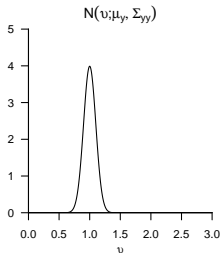
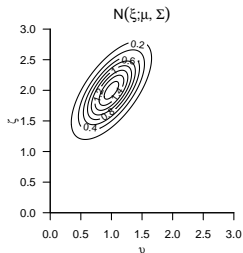
$$y \sim N(\mu_y, \Sigma_{yy}) \text{ and } z \sim N(\mu_z, \Sigma_{zz}), \quad (24)$$

Bemerkungen

- Wir verzichten auf einen Beweis
- Die Marginalverteilungen einer multivariaten Normalverteilung sind auch Normalverteilungen, insbesondere sind die eindimensionalen Marginalverteilungen einer multivariaten Normalverteilung univariate Normalverteilungen.
- Die Parameter der Marginalverteilungen ergeben sich aus den Parametern der gemeinsamen Verteilung.

Marginale Normalverteilungen

$$n := 2, k = 1, l = 1, \mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \Sigma := \begin{pmatrix} 0.10 & 0.08 \\ 0.08 & 0.15 \end{pmatrix}$$



Theorem (Gemeinsame Normalverteilungen)

x sei ein m -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_x : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, \xi \mapsto p_x(\xi) := N(\xi; \mu_x, \Sigma_{xx}) \text{ mit } \mu_x \in \mathbb{R}^m, \Sigma_{xx} \in \mathbb{R}^{m \times m}, \quad (25)$$

$A \in \mathbb{R}^{n \times m}$ sei eine Matrix, $b \in \mathbb{R}^n$ sei ein Vektor und y sei ein n -dimensionaler bedingt normalverteilter Zufallsvektor mit bedingter WDF

$$p_{y|x}(\cdot|\xi) : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, y \mapsto p_{y|x}(y|\xi) := N(y; AX + b, \Sigma_{yy}) \text{ mit } \Sigma_{yy} \in \mathbb{R}^{n \times n}. \quad (26)$$

Dann ist der $m + n$ -dimensionale Zufallsvektor (x, y) normalverteilt mit (gemeinsamer) WDF

$$p_{x,y} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, (\xi, v) \mapsto p_{x,y}(\xi, v) = N\left(\begin{pmatrix} \xi \\ v \end{pmatrix}; \mu_{x,y}, \Sigma_{x,y}\right), \quad (27)$$

mit $\mu_{x,y} \in \mathbb{R}^{m+n}$ and $\Sigma_{x,y} \in \mathbb{R}^{(m+n) \times (m+n)}$ und insbesondere

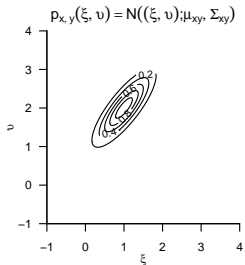
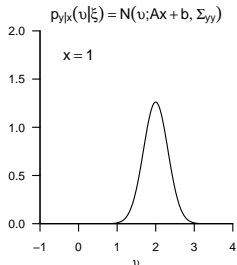
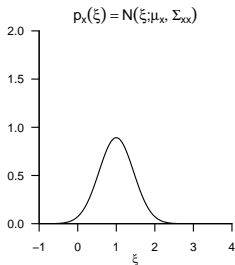
$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ A\mu_x + b \end{pmatrix} \text{ und } \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx}A^T \\ A\Sigma_{xx} & \Sigma_{yy} + A\Sigma_{xx}A^T \end{pmatrix}. \quad (28)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Eine marginale und eine bedingte multivariate Normalverteilung induzieren eine gemeinsame Normalverteilung.
- Die Parameter der gemeinsamen Verteilungen ergeben sich als linear-affine Transformation der Parameter der induzierenden Verteilungen.

Gemeinsame Verteilungen

$$m := 1, n := 1, \mu_x := 1, \Sigma_{xx} := 0.2, A := 1, b := 1, \Sigma_{yy} := 0.1$$



Theorem (Bedingte Normalverteilungen)

(x, y) sei ein $m + n$ -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_{x,y} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, (\xi, \nu) \mapsto p_{x,y}(\xi, \nu) := N\left((\xi, \nu); \mu_{x,y}, \Sigma_{x,y}\right), \quad (29)$$

mit

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \quad (30)$$

mit $x, \mu_x \in \mathbb{R}^m, y, \mu_y \in \mathbb{R}^n$ and $\Sigma_{xx} \in \mathbb{R}^{m \times m}, \Sigma_{xy} \in \mathbb{R}^{m \times n}, \Sigma_{yy} \in \mathbb{R}^{n \times n}$. Dann ist die bedingte Verteilung von x gegeben y eine m -dimensionale Normalverteilung mit bedingter WDF

$$p_{x|y}(\cdot|\nu) : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_{x|y}(\xi|\nu) := N(\xi; \mu_{x|y}, \Sigma_{x|y}) \quad (31)$$

mit

$$\mu_{x|y} = \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (Y - \mu_y) \in \mathbb{R}^m \quad (32)$$

und

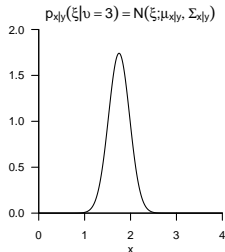
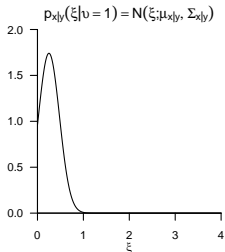
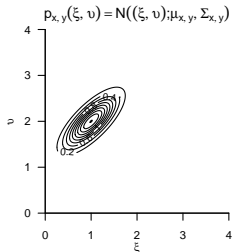
$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \in \mathbb{R}^{m \times m}. \quad (33)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die Parameter einer bedingten (multivariaten) Normalverteilung ergeben sich aus den Parameter einer gemeinsamen multivariaten Normalverteilung. Im Zusammenspiel mit den vorherigen Theoremen können die Parameter bedingter und marginale Normalverteilungen aus den Parametern der komplementären bedingten und marginalen Normalverteilungen bestimmt werden.

Bedingte Normalverteilungen

$$m := 2, n := 1, \mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \Sigma := \begin{pmatrix} 0.12 & 0.09 \\ 0.09 & 0.12 \end{pmatrix}$$



Definition

Transformationen

Unabhängigkeit

Gemeinsame Verteilungen

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie die WDF eines multivariaten normalverteilten Zufallsvektors wieder und erläutern Sie diese.
2. Geben Sie die Definitionen des Erwartungswerts und der Kovarianzmatrix eines Zufallsvektors wieder.
3. Was repräsentieren die Elemente einer Kovarianzmatrix eines Zufallsvektors?
4. Welche Werte haben der Erwartungswert und die Kovarianzmatrix eines normalverteilten Zufallsvektors?
5. Beweisen Sie die letzte Gleichung im Beweis der funktionalen Form der WDF der bivariaten Normalverteilung.
6. Visualisieren Sie die WDF eines 2-dimensionalen normalverteilten Zufallsvektors mit den Parameterwerten

$$\mu := \begin{pmatrix} 10 \\ 15 \end{pmatrix} \text{ und } \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \quad (34)$$

Generieren Sie 100 Realisierungen aus dieser Verteilung und visualisieren Sie diese.

7. Geben Sie das Theorem zur nichtsingulären linearen Transformation multivariater Normalverteilungen wieder.
8. Geben Sie das Theorem zur linear-affinen Transformation multivariater Normalverteilungen wieder.
9. Geben Sie das Theorem zu Unabhängigen normalverteilten Zufallsvariablen wieder.
10. Erläutern Sie den Begriff des sphärischen Kovarianzmatrixparameters.
11. Skizzieren Sie den Beweis des Theorem zu Unabhängigen normalverteilten Zufallsvariablen.

References

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



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

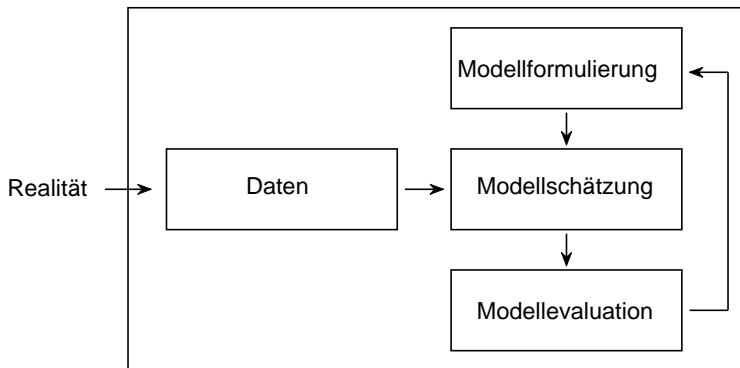
Prof. Dr. Dirk Ostwald

(5) Modellformulierung

Überblick

Datum	Einheit	Thema
08.04.2022	Grundlagen Osterpause	(1) Regression
22.04.2022	Grundlagen	(2) Korrelation
29.04.2022	Grundlagen	(3) Matrizen
06.05.2022	Grundlagen	(4) Normalverteilungen
13.05.2022	Theorie	(5) Modellformulierung
20.05.2022	Theorie	(6) Modellschätzung
27.05.2022	Theorie	(7) Modellevaluation
03.06.2021	Anwendung	(8) Studiendesign
10.06.2021	Anwendung	(9) T-Tests
17.06.2021	Anwendung	(10) Einfaktorielle Varianzanalyse
24.06.2022	Anwendung	(11) Zweifaktorielle Varianzanalyse
01.07.2022	Anwendung	(12) Multiple Regression
08.07.2022	Anwendung	(13) Kovarianzanalyse
Juli 2022	Klausurtermin	
März 2023	Klausurwiederholungstermin	

Naturwissenschaft



Modellformulierung

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (1)$$

Modellschätzung

$$\hat{\beta} = (X^T X)^{-1} X^T y, \hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (2)$$

Modellevaluation

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, F = \frac{(\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon})/p_2}{\hat{\varepsilon}^T \hat{\varepsilon}/(n - p)} \quad (3)$$

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für die wahren, aber unbekanntem, Parameterwerte (oder eine Funktion derer) abzugeben, typischerweise basierend auf der Beobachtung einer Datenrealisierung.

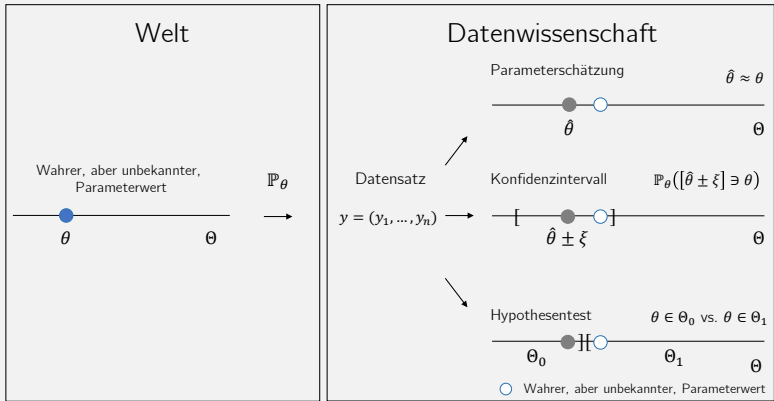
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst sinnvollen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert, sich in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Modell und Standardprobleme der Frequentistischen Inferenz



Standardannahmen Frequentistischer Inferenz

Gegeben sei ein statistisches Modell mit. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. Aus Frequentistischer Sicht kann man unendlich oft Datensätze basierend auf einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right)^T \text{ mit } \hat{\beta}^{(1)} = (X^T X)^{-1} X^T y^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right)^T \text{ mit } \hat{\beta}^{(2)} = (X^T X)^{-1} X^T y^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right)^T \text{ mit } \hat{\beta}^{(3)} = (X^T X)^{-1} X^T y^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right)^T \text{ mit } \hat{\beta}^{(4)} = (X^T X)^{-1} X^T y^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung. Was zum Beispiel ist die Verteilung von $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \hat{\beta}^{(3)}, \hat{\beta}^{(4)}, \dots$ also die Verteilung der Zufallsvariable $\hat{\beta} := (X^T X)^{-1} X^T y$? Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Anwendungsbeispiele Einheit (5) - (7)

- Unabhängig und identisch normalverteilte Zufallsvariablen | Einstichproben-T-Test
- Einfache lineare Regression

Anwendungsbeispiele Einheit (8) - (13)

- Zweistichproben-T-Tests
- Einfaktorielle Varianzanalyse
- Zweifaktorielle Varianzanalyse
- Multiple Regression
- Kovarianzanalyse

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Definition (Allgemeines Lineares Modell)

Es sei

$$y = X\beta + \varepsilon, \quad (4)$$

wobei

- y ein n -dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird,
- $X \in \mathbb{R}^{n \times p}$ eine vorgegebene Matrix ist, die *Designmatrix* genannt wird,
- $\beta \in \mathbb{R}^p$ ein unbekannter Parametervektor ist, der *Betaparametervektor* genannt wird und
- ε ein n -dimensionaler nicht-beobachtbarer Zufallsvektor ist, der *Zufallsfehler* genannt wird und für den angenommen wird, dass mit einem unbekanntem Varianzparameter $\sigma^2 > 0$ gilt, dass

$$\varepsilon \sim N(0_n, \sigma^2 I_n). \quad (5)$$

Dann wird (4) *Allgemeines Lineares Modell (ALM) in generativer Form* genannt.

Bemerkungen

- y ist ein Zufallsvektor, weil er aus der Addition des Zufallsvektors ε zu dem Vektor $X\beta \in \mathbb{R}^n$ resultiert.
- Wir nennen $X\beta \in \mathbb{R}^n$ den *deterministischen Modellaspekt* und ε den *probabilistischen Modellaspekt*.
- $n \in \mathbb{N}$ bezeichnet durchgängig die Anzahl an Datenpunkten.
- $p \in \mathbb{N}$ bezeichnet durchgängig die Anzahl an Betaparametern.
- Die Gesamtzahl an Parametern des ALMs ist $p + 1$ (p Betaparameter und 1 Varianzparameter).
- Der Betaparametervektor wird auch *Gewichtsvektor* oder *Effektvektor* genannt.
- Weil der Kovarianzmatrixparameter von ε als sphärisch angenommen wird, sind die $\varepsilon_1, \dots, \varepsilon_n$ unabhängige normalverteilte Zufallsvariablen mit identischem Varianzparameter; weil zusätzlich der Erwartungswertparameter von ε als 0_n angenommen wird, sind die $\varepsilon_1, \dots, \varepsilon_n$ auch identisch normalverteilte Zufallsvariablen.
- Für jede Komponente $y_i, i = 1, \dots, n$ von y impliziert (4) nach Definition des Matrixprodukts, dass

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2), \quad (6)$$

wobei $x_{ij} \in \mathbb{R}$ das ij te Element der Designmatrix X bezeichnet.

Theorem (ALM Datenverteilung)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7)$$

das ALM in generativer Form. Dann gilt

$$y \sim N(X\beta, \sigma^2 I_n). \quad (8)$$

Beweis

Mit dem Theorem zur linear-affinen Transformation multivariater Normalverteilungen gilt für $\varepsilon \sim N(0_n, \sigma^2 I_n)$ und $y := I_n \varepsilon + X\beta$, dass

$$y \sim N \left(I_n 0_n + X\beta, I_n (\sigma^2 I_n) I_n^T \right) = N(X\beta, \sigma^2 I_n). \quad (9)$$

Bemerkungen

- Im ALM sind die Daten y also ein n -dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter $X\beta \in \mathbb{R}^n$ und Kovarianzmatrixparameter $\sigma^2 I_n \in \mathbb{R}^{n \times n}$.
- Die Komponenten y_1, \dots, y_n von y , also die Datenpunkte, sind damit unabhängige, aber im Allgemeinen nicht identisch verteilte, normalverteilte Zufallsvariablen der Form $y_i \sim N \left((X\beta)_i, \sigma^2 \right)$ für $i = 1, \dots, n$.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Unabhängige und identisch normalverteilte Zufallsvariablen

Wir betrachten das Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2 ,

$$y_i \sim N(\mu, \sigma^2) \text{ für } i = 1, \dots, n. \quad (10)$$

Dann gilt, dass (10) äquivalent ist zu

$$y_i = \mu + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \text{ mit unabhängigen } \varepsilon_i. \quad (11)$$

In Matrixschreibweise ist dies wiederum äquivalent zu

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}^1, \sigma^2 > 0. \quad (12)$$

Bemerkungen

- Wir kennen dieses Modell bereits aus Einheit (9) Grundbegriffe Frequentistischer Inferenz in Wahrscheinlichkeitstheorie und Frequentistische Inferenz, dort haben wir es geschrieben als

$$X_1, \dots, X_n \sim N(\mu, \sigma^2) \text{ mit } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}. \quad (13)$$

- Bitte verwechseln Sie nicht die Designmatrix $X \in \mathbb{R}^{n \times p}$ des ALMs und die Zufallsvariablen X_1, \dots, X_n der Szenarien in Wahrscheinlichkeitstheorie und Frequentistische Inferenz.

Unabhängige und identisch normalverteilte Zufallsvariablen

```
# Libraries
library(MASS) # Multivariate Normalverteilung

# Modellformulierung
n = 12 # Anzahl von Datenpunkten
p = 1 # Anzahl von Betaparameter
X = matrix(rep(1,n), nrow = n) # Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = 2 # wahrer, aber unbekannter, Betaparameter
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
print(y)

> [1] 2.629 1.446 1.717 1.756 1.753 0.178 3.148 2.622 1.994
> [10] -0.437 2.255 0.600
```

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Einfache lineare Regression

Wir betrachten das generative Modell der einfachen linearen Regression aus Einheit (1) Regression,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n, \quad (14)$$

wobei wir hier nun die Zufallsvariablen y_1, \dots, y_n mit kleinen Buchstaben bezeichnen.

Wir haben bereits gesehen, dass dieses Modell äquivalent ist zu dem Normalverteilungsmodell der Regression

$$y_i \sim N(\mu_i, \sigma^2) \text{ mit } \mu_i := \beta_0 + \beta_1 x_i \text{ für } i = 1, \dots, n. \quad (15)$$

In Matrixschreibweise ist dies wiederum äquivalent zu

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^2, \sigma^2 > 0. \quad (16)$$

Einfache lineare Regression

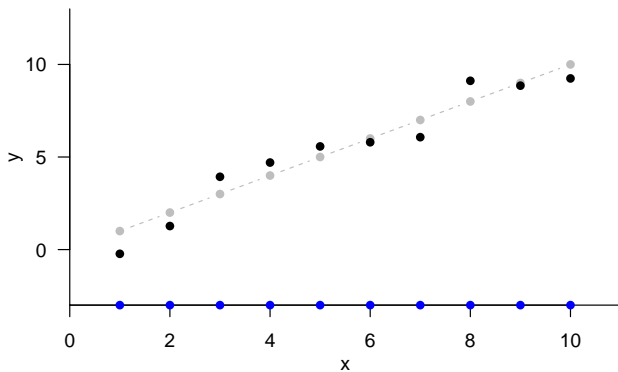
```
# Libraries
library(MASS) # Multivariate Normalverteilung

# Modellformulierung
n = 10 # Anzahl von Datenpunkten
p = 2 # Anzahl von Betaparametern
x = 1:n # Prädiktorwerte
X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter, Betaparameter
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
print(y)

> [1] 1.36 2.47 2.09 4.54 4.95 5.48 5.14 8.51 7.37 12.07
```

Einfache lineare Regression



• x_i • $X\beta$ für $\beta_0 := 0, \beta_1 := 1$ • (x_i, y_i)

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie das naturwissenschaftliche Paradigma.
2. Erläutern Sie die Standardprobleme der Frequentistischen Inferenz.
3. Setzen Sie das naturwissenschaftliche Paradigma und die Frequentistische Inferenz in Beziehung.
4. Geben Sie die Definition des ALMs in generativer Form wieder.
5. Erläutern Sie die deterministischen und probabilistischen Aspekte des ALMs.
6. Wieviele Parameter hat das ALM mit sphärischer Kovarianzmatrix?
7. Warum sind die Komponenten des ALM Zufallsfehler unabhängig und identisch verteilt?
8. Geben Sie das Theorem zur ALM Datenverteilung wieder.
9. Sind die Komponenten des ALM Datenvektors unabhängig und identisch verteilt?
10. Schreiben Sie das Szenario n unabhängiger und identisch verteilter Zufallsvariablen als ALM in Matrixschreibweise.
11. Schreiben Sie das Szenario der einfachen linearen Regression als ALM in Matrixschreibweise.
12. Generieren Sie 100 Datensätze von 12 unabhängigen und identisch verteilten Zufallsvariablen.
13. Generieren Sie 100 Datensätze von einem einfachen linearen Regressionsmodell mit 12 äquidistanten Werten der unabhängigen Variable im Intervall $[1, 2]$, wobei $x_1 := 1$ und $x_{12} := 2$ sein sollen.



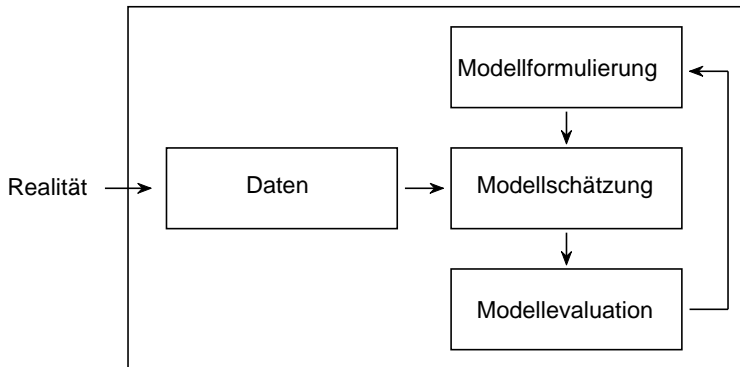
Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(6) Modellschätzung

Naturwissenschaft



Modellformulierung

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (1)$$

Modellschätzung

$$\hat{\beta} = (X^T X)^{-1} X^T y, \hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (2)$$

Modellevaluation

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, F = \frac{(\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon})/p_2}{\hat{\varepsilon}^T \hat{\varepsilon}/(n - p)} \quad (3)$$

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für die wahren, aber unbekanntem, Parameterwerte (oder eine Funktion derer) abzugeben, typischerweise basierend auf der Beobachtung einer Datenrealisierung.

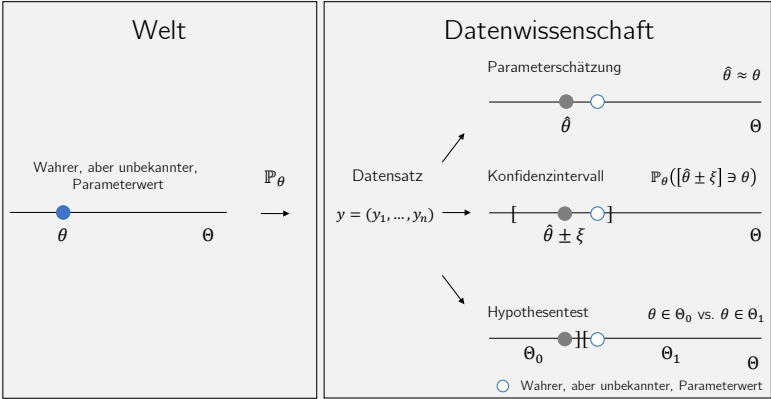
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst sinnvollen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert, sich in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Modell und Standardprobleme der Frequentistischen Inferenz



Standardannahmen Frequentistischer Inferenz

Gegeben sei ein statistisches Modell mit. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. Aus Frequentistischer Sicht kann man unendlich oft Datensätze basierend auf einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right)^T \text{ mit } \hat{\beta}^{(1)} = (X^T X)^{-1} X^T y^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right)^T \text{ mit } \hat{\beta}^{(2)} = (X^T X)^{-1} X^T y^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right)^T \text{ mit } \hat{\beta}^{(3)} = (X^T X)^{-1} X^T y^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right)^T \text{ mit } \hat{\beta}^{(4)} = (X^T X)^{-1} X^T y^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung. Was zum Beispiel ist die Verteilung von $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \hat{\beta}^{(3)}, \hat{\beta}^{(4)}, \dots$ also die Verteilung der Zufallsvariable $\hat{\beta} := (X^T X)^{-1} X^T y$? Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Theorem (Betaparameterschätzer)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (4)$$

das ALM in generativer Form und es sei

$$\hat{\beta} := (X^T X)^{-1} X^T y. \quad (5)$$

der *Betaparameterschätzer*. Dann gilt, dass

$$\hat{\beta} = \arg \min_{\tilde{\beta}} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (6)$$

und dass $\hat{\beta}$ ein unverzerrter Maximum Likelihood Schätzer von $\beta \in \mathbb{R}^p$ ist.

Bemerkungen

- Das Theorem gibt eine Formel an, um β anhand von Designmatrix und Daten zu schätzen.
- Als ML Schätzer ist $\hat{\beta}$ konsistent, asymptotisch normalverteilt und asymptotisch effizient.
- Wir sehen später, dass $\hat{\beta}$ sogar normalverteilt ist.
- Außerdem hat $\hat{\beta}$ die "kleinste Varianz" in der Klasse der linearen unverzerrten Schätzer von β .
- Letztere Eigenschaft ist Kernaussage des *Gauss-Markov Theorems*, auf das wir hier nicht näher eingehen wollen.
- Für eine Diskussion und einen Beweis des Gauss-Markov Theorems siehe z.B. Searle (1971), Kapitel 3.

Beweis

(1) Wir zeigen in einem ersten Schritt, dass $\hat{\beta}$ die Summe der Abweichungsquadrate

$$(y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (7)$$

minimiert. Dazu halten wir zunächst fest, dass

$$\hat{\beta} = (X^T X)^{-1} X^T y \Leftrightarrow X^T X \hat{\beta} = X^T y \Leftrightarrow X^T y - X^T X \hat{\beta} = 0_p \Leftrightarrow X^T (y - X \hat{\beta}) = 0_p. \quad (8)$$

Weiterhin gilt dann auch, dass

$$X^T (y - X \hat{\beta}) = 0_p \Leftrightarrow (X^T (y - X \hat{\beta}))^T = 0_p^T \Leftrightarrow (y - X \hat{\beta})^T X = 0_p^T \quad (9)$$

Weiterhin halten wir ohne Beweis fest, dass für jede Matrix $X \in \mathbb{R}^{n \times p}$ gilt, dass

$$z^T X^T X z \geq 0 \text{ für alle } z \in \mathbb{R}^p. \quad (10)$$

Wir betrachten nun für festes y und ein beliebiges $\tilde{\beta}$ die Summe der Abweichungsquadrate

$$(y - X\tilde{\beta})^T (y - X\tilde{\beta}). \quad (11)$$

Beweis (fortgeführt)

Es ergibt sich

$$\begin{aligned}(y - X\tilde{\beta})^T (y - X\tilde{\beta}) &= (y - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta})^T (y - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta}) \\ &= ((y - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta}))^T ((y - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta})) \\ &= (y - X\hat{\beta})^T (y - X\hat{\beta}) + (y - X\hat{\beta})^T X(\hat{\beta} - \tilde{\beta}) \\ &\quad + (\hat{\beta} - \tilde{\beta})^T X^T (y - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \quad (12) \\ &= (y - X\hat{\beta})^T (y - X\hat{\beta}) + 0_p^T (\hat{\beta} - \tilde{\beta}) \\ &\quad + (\hat{\beta} - \tilde{\beta})^T 0_p + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\ &= (y - X\hat{\beta})^T (y - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta})\end{aligned}$$

Auf der rechten Seite obiger Gleichung ist nur der zweite Term von $\tilde{\beta}$ abhängig. Da für diesen Term gilt, dass

$$(\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \geq 0 \quad (13)$$

nimmt dieser Term genau dann seinen Minimalwert 0 an, wenn

$$(\hat{\beta} - \tilde{\beta}) = 0_p \Leftrightarrow \tilde{\beta} = \hat{\beta}. \quad (14)$$

Also gilt

$$\hat{\beta} = \arg \min_{\tilde{\beta}} (y - X\tilde{\beta})^T (y - X\tilde{\beta}). \quad (15)$$

Beweis (fortgeführt)

(2) Um zu zeigen, dass $\hat{\beta}$ ein Maximum Likelihood Schätzer ist, betrachten wir für festes $y \in \mathbb{R}^n$ und festes $\sigma^2 > 0$ die Log-Likelihood Funktion

$$\ell : \mathbb{R}^p \rightarrow \mathbb{R}, \tilde{\beta} \mapsto \ln p_{\tilde{\beta}}(y) = \ln N(y; X\tilde{\beta}, \sigma^2 I_n) \quad (16)$$

wobei gilt, dass

$$\begin{aligned} \ln N(y; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left((2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \right) \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta}). \end{aligned} \quad (17)$$

Dabei hängt allein der Term $-\frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta})$ von $\tilde{\beta}$ ab. Weil aber $(y - X\tilde{\beta})^T (y - X\tilde{\beta}) \geq 0$, gilt wird dieser Term aufgrund des negativen Vorzeichen maximal, wenn $(y - X\tilde{\beta})^T (y - X\tilde{\beta})$ minimal wird. Dies ist aber wie oben gezeigt genau für $\tilde{\beta} = \hat{\beta}$ der Fall.

(3) Die Unverzerrtheit von $\hat{\beta}$ schließlich ergibt sich aus

$$\mathbb{E}(\hat{\beta}) = \mathbb{E} \left((X^T X)^{-1} X^T y \right) = (X^T X)^{-1} X^T \mathbb{E}(y) = (X^T X)^{-1} X^T X \beta = \beta. \quad (18)$$

Theorem (Varianzparameterschätzer)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (19)$$

das ALM in generativer Form. Dann ist

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (20)$$

ein unverzerrter Schätzer von $\sigma^2 > 0$.

Bemerkungen

- Es handelt sich bei $\hat{\sigma}^2$ *nicht* um einen ML Schätzer von σ^2 .
- Für einen Beweis siehe z.B. Searle (1971), Kapitel 3 oder Rencher and Schaalje (2008), Kapitel 7.
- Mit Definition des *Residuenvektors* und der *Residuen*

$$\hat{\varepsilon} := y - X\hat{\beta} \text{ und } \hat{\varepsilon}_i := y_i - (X\hat{\beta})_i, i = 1, \dots, n, \quad (21)$$

respektive, bieten sich für $\hat{\sigma}^2$ auch folgende Schreibweisen an:

$$\hat{\sigma}^2 = \frac{1}{n - p} \hat{\varepsilon}^T \hat{\varepsilon} = \frac{1}{n - p} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n - p} \sum_{i=1}^n (y_i - (X\hat{\beta})_i)^2 \quad (22)$$

- σ^2 wird also durch die summierten quadrierten Residuen geschätzt.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Unabhängige und identisch normalverteilte Zufallsvariablen

Wir betrachten das Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2 ,

$$y_i \sim N(\mu, \sigma^2) \text{ für } i = 1, \dots, n. \quad (23)$$

Dann gilt, wie unten gezeigt,

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y} \text{ und } \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2. \quad (24)$$

In diesem Fall ist also der Betaparameterschätzer mit dem Stichprobenmittel \bar{y} der y_1, \dots, y_n und der Varianzparameterschätzer mit der Stichprobenvarianz s_y^2 der y_1, \dots, y_n identisch.

Unabhängige und identisch normalverteilte Zufallsvariablen

Für $\hat{\beta}$ ergibt sich

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left(\mathbf{1}_n^T \mathbf{1}_n \right)^{-1} \mathbf{1}_n^T y \\ &= \left(\begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= n^{-1} \sum_{i=1}^n y_i \\ &= \frac{1}{n} \sum_{i=1}^n y_i \\ &=: \bar{y}.\end{aligned}$$

Unabhängige und identisch normalverteilte Zufallsvariablen

Für $\hat{\sigma}^2$ ergibt sich

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-1} (y - X\hat{\beta})^T (y - X\hat{\beta}) \\ &= \frac{1}{n-1} (y - 1_n \bar{y})^T (y - 1_n \bar{y}) \\ &= \frac{1}{n-1} \left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y} \right)^T \left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y} \right) \\ &= \frac{1}{n-1} (y_1 - \bar{y} \quad \cdots \quad y_n - \bar{y}) \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \\ &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &=: s_y^2.\end{aligned}$$

Unabhängige und identisch normalverteilte Zufallsvariablen

Parameterschätzung

```
# Libraries
library(MASS) # Multivariate Normalverteilung

# Modellformulierung
n = 12 # Anzahl Datenpunkte
p = 1 # Anzahl Betaparameter
X = matrix(rep(1,n), nrow = n) # Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = 2 # wahrer, aber unbekannter, Betaparameter
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %>% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs

# Parameterschätzung
beta_hat = solve(t(X) %>% X) %>% t(X) %>% y # Betaparameterschätzer
eps_hat = y - X %>% beta_hat # Residuenvektor
sigsqr_hat = (t(eps_hat) %>% eps_hat) / (n-p) # Varianzparameterschätzer

# Ausgabe
cat("beta : ", beta,
    "\nhat{beta} : ", beta_hat,
    "\nsigsqr : ", sigsqr,
    "\nhat{sigsqr}: ", sigsqr_hat)

> beta : 2
> hat{beta} : 1.64
> sigsqr : 1
> hat{sigsqr}: 1.12
```


Unabhängige und identisch normalverteilte Zufallsvariablen

Simulation der Schätzerunverzerrtheit

```
# Libraries
library(MASS)

# Modellformulierung
n = 12
p = 1
X = matrix(rep(1,n), nrow = n)
I_n = diag(n)
beta = 2
sigsqr = 1

# Multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter, Betaparameter
# wahrer, aber unbekannter, Varianzparameter

# Frequentistische Simulation
nsim = 1e4
beta_hat = rep(NA,n)
sigsqr_hat = rep(NA,n)
for(i in 1:nsim){
  y = mvrnorm(1, X %*% beta, sigsqr*I_n)
  beta_hat[i] = solve(t(X) %*% X) %*% t(X) %*% y
  eps_hat = y - X %*% beta_hat[i]
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat) / (n-p)
}

# Ausgabe
cat("Wahrer, aber unbekannter, Betaparameter : ", beta,
    "\nGeschätzter Erwartungswert des Betaparameterschätzers : ", mean(beta_hat),
    "\nWahrer, aber unbekannter, Varianzparameter : ", sigsqr,
    "\nGeschätzter Erwartungswert des Varianzparameterschätzers : ", mean(sigsqr_hat))

> Wahrer, aber unbekannter, Betaparameter : 2
> Geschätzter Erwartungswert des Betaparameterschätzers : 2
> Wahrer, aber unbekannter, Varianzparameter : 1
> Geschätzter Erwartungswert des Varianzparameterschätzers : 0.998
```

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Selbstkontrollfragen

Einfache lineare Regression

Wir betrachten das generative Modell der einfachen linearen Regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n, \quad (25)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix} \text{ and } \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \quad (26)$$

wobei \bar{x} und \bar{y} die Stichprobenmittel der x_1, \dots, x_n und y_1, \dots, y_n , respektive, bezeichnen, c_{xy} die Stichprobenkovarianz der x_1, \dots, x_n und s_x^2 die Stichprobenvarianz der x_1, \dots, x_n bezeichnen, wie unten gezeigt.

Wir halten fest, dass für eine parametrische Designmatrixspalte sich der entsprechende Betaparameterschätzer aus der Stichprobenkovarianz der respektiven Spalte mit den Daten geteilt durch die Stichprobenvarianz der entsprechenden Spalte ergibt und somit einer "standardisierten" Stichprobenkovarianz entspricht. Wie in (1) Regression sind die Bezeichnungen "Stichproben"kovarianz und "Stichproben"varianz bezüglich der x_1, \dots, x_n hier lediglich formal gemeint, da keine Annahme zugrundeliegt, dass die x_1, \dots, x_n Realisierungen von Zufallsvariablen sind. Die x_1, \dots, x_n sind vorgegebene Werte. Ein Vergleich mit den Parametern der Ausgleichsgerade in (1) Regression zeigt weiterhin die Identität der Betaparameterschätzerkomponenten $\hat{\beta}_0$ und $\hat{\beta}_1$ mit den dort unter dem Kriterium der Minimierung der quadrierten vertikalen Abweichungen hergeleiteten Parametern. Dies überrascht nicht, da sowohl $\hat{\beta}$ als auch die Parameter der Ausgleichsgerade den Wert

$$q(\tilde{\beta}) = \sum_{i=1}^n (y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2 = (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (27)$$

hinsichtlich $\tilde{\beta}$ minimieren.

Einfache lineare Regression

Um die Form des Betaparameterschätzers herzuleiten, halten wir zunächst fest, dass

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \bar{y} - \sum_{i=1}^n \bar{x} y_i + \sum_{i=1}^n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \bar{y} n \bar{x} - \bar{x} n \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y},\end{aligned}\tag{28}$$

Einfache lineare Regression

Weiterhin halten wir fest, dass

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2x_i\bar{x} + \sum_{i=1}^n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2.\end{aligned}\tag{29}$$

Einfache lineare Regression

Aus der Definition von $\hat{\beta}$ ergibt sich

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left(\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \left(\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \right)^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\ &= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.\end{aligned}\tag{30}$$

Die Inverse von $X^T X$ ist gegeben durch

$$\frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},\tag{31}$$

weil

$$\begin{aligned} & \frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} \frac{ns_x^2}{n} + n\bar{x}^2 - n\bar{x}^2 & \frac{s_x^2 n\bar{x}}{n} + n\bar{x}^2 \bar{x} - \bar{x} \sum_{i=1}^n x_i^2 \\ -\bar{x}n + n\bar{x} & -n\bar{x}^2 + \sum_{i=1}^n x_i^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & s_x^2 \bar{x} - \bar{x} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\ 0 & \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & s_x^2 \bar{x} - \bar{x} s_x^2 \\ 0 & s_x^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & 0 \\ 0 & s_x^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{32}$$

Einfache lineare Regression

Es ergibt sich also

$$\begin{aligned}\hat{\beta} &= \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_x^2} & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2} \right) n\bar{y} - \frac{\bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i}{s_x^2} - \frac{n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{n\bar{y}}{n} + \frac{\bar{x}^2 n\bar{y}}{s_x^2} - \frac{\bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} + \frac{\bar{x} n\bar{x}\bar{y} - \bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} - \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \bar{x} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix}.\end{aligned}\tag{33}$$

Einfache lineare Regression

Parameterschätzung

```
# Libraries
library(MASS) # Multivariate Normalverteilung
library(matlib) # Matrizenrechnung

# Modellformulierung
n = 10 # Anzahl Datenpunkte
p = 2 # Anzahl Betaparameter
x = 1:n # Prädiktorwerte
X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter, Betaparameter
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs

# Parameterschätzung
beta_hat = inv(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer

# Ausgabe
cat("beta : ", beta,
    "\nhat{beta} : ", beta_hat,
    "\nsigsqr : ", sigsqr,
    "\nhat{sigsqr}: ", sigsqr_hat)

> beta : 0 1
> hat{beta} : -1.04 1.07
> sigsqr : 1
> hat{sigsqr}: 0.337
```

Einfache lineare Regression

Simulation der Schätzerunverzerrtheit

```
# Libraries
library(MASS)
library(matlib)

# Modellformulierung
n      = 10
p      = 2
x      = 1:n
X      = matrix(c(rep(1,n),x), nrow = n)
I_n    = diag(n)
beta   = matrix(c(0,1), nrow = p)
sigsqr = 1

# Frequentistische Simulation
nsim   = 1e4
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)
sigsqr_hat = rep(NaN,nsim)
for(i in 1:nsim){
  y      = mvrnorm(1, X %*% beta, sigsqr*I_n)
  beta_hat[,i] = inv(t(X) %*% X) %*% t(X) %*% y
  eps_hat = y - X %*% beta_hat[,i]
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat) / (n-p)
}

# Ausgabe
cat("Wahrer, aber unbekannter, Betaparameter      : ", beta,
    "\nGeschätzter Erwartungswert des Betaparameterschätzers : ", rowMeans(beta_hat),
    "\nWahrer, aber unbekannter, Varianzparameter      : ", sigsqr,
    "\nGeschätzter Erwartungswert des Varianzparameterschätzers : ", mean(sigsqr_hat))

> Wahrer, aber unbekannter, Betaparameter      : 0 1
> Geschätzter Erwartungswert des Betaparameterschätzers : -0.00191 1
> Wahrer, aber unbekannter, Varianzparameter      : 1
> Geschätzter Erwartungswert des Varianzparameterschätzers : 1.01
```

```
# Multivariate Normalverteilung
```

```
# Matrizenrechnung
```

```
# Anzahl Datenpunkte
```

```
# Anzahl Betaparameter
```

```
# Prädiktorwerte
```

```
# Designmatrix
```

```
# n x n Einheitsmatrix
```

```
# wahrer, aber unbekannter, Betaparameter
```

```
# wahrer, aber unbekannter, Varianzparameter
```

```
# Anzahl Realisierungen des n-dimensionalen ZVs
```

```
#  $\hat{\beta}$  Realisierungsarray
```

```
#  $\hat{\text{sigsqr}}$  Realisierungsarray
```

```
# Simulationsiterationen
```

```
# Datenrealisierung
```

```
# Betaparameterschätzer
```

```
# Residuenvektor
```

```
# Varianzparameterschätzer
```

Selbstkontrollfragen

1. Geben Sie das Betaparameterschätzer Theorem wieder.
2. Geben Sie das Varianzparameterschätzer Theorem wieder.
3. Geben Sie die Beta- und Varianzparameterschätzer des ALM Szenarios von n unabhängigen und identisch normalverteilten Zufallsvariablen wieder.
4. Geben Sie die Beta- und Varianzparameterschätzer des ALM Szenarios der einfachen linearen Regression wieder.
5. Wie unterscheiden sich die Betaparameterschätzer des ALM Szenarios der einfachen linearen Regression und die Parameter der Ausgleichsgerade aus Einheit (1) Regression?
6. Simulieren Sie die Unverzerrtheit des Varianzparameterschätzers im ALM Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen in einem R Skript.
7. Simulieren Sie die Unverzerrtheit des Varianzparameterschätzers im ALM Szenario der einfachen linearen Regression in einem R Skript.

References

- Rencher, Alvin C., and G. Bruce Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. Hoboken, N.J: Wiley-Interscience.
- Searle, S. R. 1971. *Linear Models*. A Wiley Publication in Mathematical Statistics. New York: Wiley.



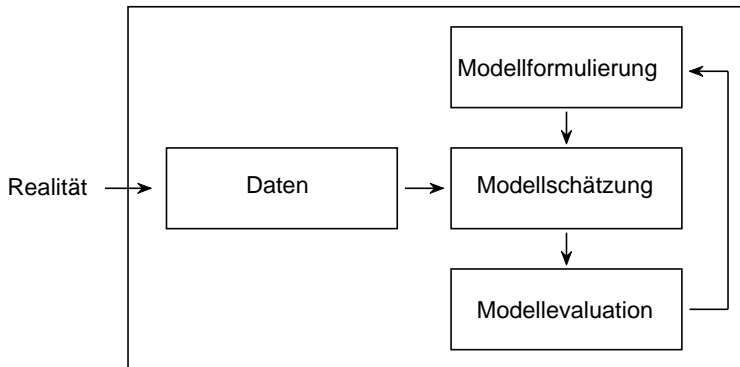
Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(7) Modellevaluation

Naturwissenschaft



Modellformulierung

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (1)$$

Modellschätzung

$$\hat{\beta} = (X^T X)^{-1} X^T y, \hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (2)$$

Modellevaluation

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, F = \frac{(\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon})/p_2}{\hat{\varepsilon}^T \hat{\varepsilon}/(n - p)} \quad (3)$$

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für die wahren, aber unbekanntem, Parameterwerte (oder eine Funktion derer) abzugeben, typischerweise basierend auf der Beobachtung einer Datenrealisierung.

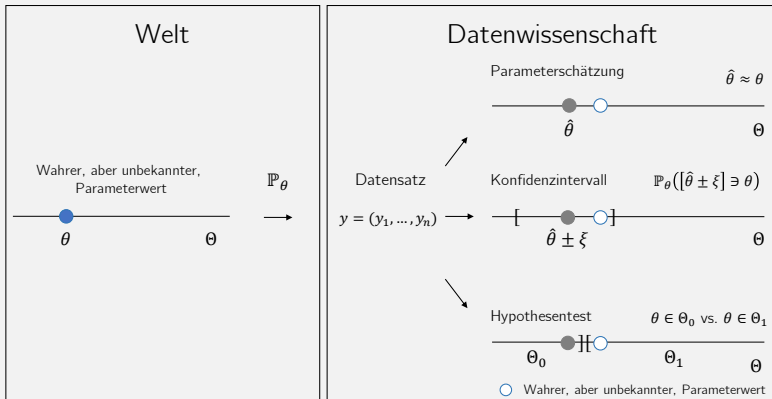
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst sinnvollen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert, sich in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Modell und Standardprobleme der Frequentistischen Inferenz



$$\theta := (\beta, \sigma^2), \Theta := \mathbb{R}^p \times \mathbb{R}_{>0} \quad \mathbb{P}_\theta(y) := \mathbb{P}_{\beta, \sigma^2}(y) \text{ mit WDF } p_{\beta, \sigma^2}(v) := N(v; X\beta, \sigma^2 I_n)$$

Standardannahmen Frequentistischer Inferenz

Gegeben sei ein statistisches Modell. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. Aus Frequentistischer Sicht kann man unendlich oft Datensätze basierend auf einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right)^T \text{ mit } \hat{\beta}^{(1)} = (X^T X)^{-1} X^T y^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right)^T \text{ mit } \hat{\beta}^{(2)} = (X^T X)^{-1} X^T y^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right)^T \text{ mit } \hat{\beta}^{(3)} = (X^T X)^{-1} X^T y^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right)^T \text{ mit } \hat{\beta}^{(4)} = (X^T X)^{-1} X^T y^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung. Was zum Beispiel ist die Verteilung von $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \hat{\beta}^{(3)}, \hat{\beta}^{(4)}, \dots$ also die Verteilung der Zufallsvariable $\hat{\beta} := (X^T X)^{-1} X^T y$? Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Überblick

- Unter den Standardannahmen der Frequentistischen Inferenz kann man fragen, wie $\hat{\beta}$ und $\hat{\sigma}^2$ verteilt sind.
- Das erste Theorem dieses Abschnitts besagt, dass $\hat{\beta}$ multivariat normalverteilt ist
- Das zweite Theorem dieses Abschnitts besagt, dass eine skalierte Version von $\hat{\sigma}^2$ χ^2 -verteilt ist.
- Diese Ergebnisse sind sehr ähnlich zu den Ergebnissen aus (7) Transformationen der Normalverteilung in WTFI.
- T- und F-Statistiken sind Funktionen von $\hat{\beta}$ und $\hat{\sigma}^2$.
- Die Verteilungen von $\hat{\beta}$ und $\hat{\sigma}^2$ bilden die Grundlage der Verteilungen der T- und F-Statistiken.
- In der konkreten Datenanalyse eines Datensatzes y berechnet man genau ein $\hat{\beta}$ und ein $\hat{\sigma}^2$.
- Probabilistische Aussagen über die aus diesen $\hat{\beta}$ und $\hat{\sigma}^2$ berechneten T- und F-Statistiken beziehen sich auf die Standardannahmen Frequentistischer Inferenz und werden vor diesem Hintergrund verstanden.

Theorem (Frequentistische Verteilung des Betaparameterschätzers)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (4)$$

das ALM in generativer Form. Weiterhin sei

$$\hat{\beta} := (X^T X)^{-1} X^T y \quad (5)$$

der Betaparameterschätzer. Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (6)$$

Bemerkungen

- Es gilt also wie bereits gesehen $\mathbb{E}(\hat{\beta}) = \beta$ und außerdem $\mathbb{C}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.
- Die Varianzen der Komponenten von $\hat{\beta}$ sind die Diagonalelemente von $\mathbb{C}(\hat{\beta})$, also

$$\mathbb{V}(\hat{\beta}_i) = (\sigma^2 (X^T X)^{-1})_{ii} \text{ für } i = 1, \dots, p. \quad (7)$$

- Die Streuung von $\hat{\beta}$ hängt von σ^2 und der Designmatrix X ab. σ^2 ist ein experimentell nicht zu beeinflussender wahrer, aber unbekannter, Parameter X dagegen kann so gewählt werden, um zum Beispiel die Diagonalelemente von $\mathbb{C}(\hat{\beta})$ bei festem σ^2 zu minimieren.

Beweis

Das Theorem folgt direkt mit dem Theorem zur linearen Transformation von multivariaten Normalverteilung aus Einheit (4) Normalverteilungen. Speziell gilt hier:

$$\hat{\beta} \sim N \left((X^T X)^{-1} X^T X \beta, (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T \right). \quad (8)$$

Dabei können der Erwartungswertparameter wie folgt

$$(X^T X)^{-1} X^T X \beta = \beta, \quad (9)$$

und der Kovarianzmatrixparameter wie folgt

$$\begin{aligned} (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T &= (X^T X)^{-1} X^T (\sigma^2 I_n) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned} \quad (10)$$

Dabei hier die erste Gleichung aus der Tatsache, dass sowohl $X^T X$ als auch ihre Inverse $(X^T X)^{-1}$ symmetrische Matrizen sind. Damit folgt dann aber sofort

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (11)$$

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \quad (12)$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen. Wir haben bereits gesehen, dass $\hat{\beta} = \bar{y}$. Das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers impliziert damit

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (13)$$

Das Stichprobenmittel von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2 ist also normalverteilt mit Erwartungswertparameter μ und Varianzparameter σ^2/n . Wir haben diese Tatsache bereits in Einheit (7) Transformationen der Normalverteilungen in Wahrscheinlichkeitstheorie und Frequentistische Inferenz unter dem Begriff der Mittelwertstransformation kennengelernt.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

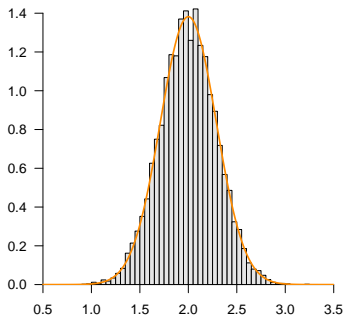
```
# Multivariate Normalverteilung
library(MASS)

# Modellformulierung
n      = 12                # Anzahl von Datenpunkten
p      = 1                # Anzahl von Betparametern
X      = matrix(rep(1,n), nrow = n) # Designmatrix
I_n    = diag(n)         # n x n Einheitsmatrix
beta   = 2               # wahrer, aber unbekannter, Betaparameter
sigsqr = 1               # wahrer, aber unbekannter, Varianzparameter

# Frequentistische Simulation
nsim   = 1e4             # Anzahl Realisierungen des n-dimensionalen ZVs
beta_hat = rep(NA,nsim) # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y      = mvrnorm(1, X %>% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
  beta_hat[i] = solve(t(X) %>% X) %>% t(X) %>% y # \hat{\beta} = (X^T X)^{-1} X^T y
}
```

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

$$N\left(\hat{\mu}; \mu, \frac{\sigma^2}{n}\right)$$



Beispiel (2) Einfache lineare Regression

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ with } \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0. \quad (14)$$

das ALM Szenario der einfachen linearen Regression. Wir haben bereits gesehen, dass

$$\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{s_{xx}} \begin{pmatrix} \frac{s_{xx}}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \text{ mit } s_{xx} := \sum_{i=1}^n (x_i - \bar{x})^2. \quad (15)$$

Die Varianz des Offsetparameterschätzers hängt also sowohl von der Summe der quadrierten Differenzen und dem Stichprobenmittel der unabhängigen Variablen x_1, \dots, x_n ab, wohingegen die Varianz des Steigungsparameterschätzers nur von der Summe der quadrierten Differenzen der x_1, \dots, x_n abhängt. Die Kovarianz von Offset- und Steigungsparameterschätzern hängt vom Mittelwert der x_1, \dots, x_n ab.

Beispiel (2) Einfache lineare Regression

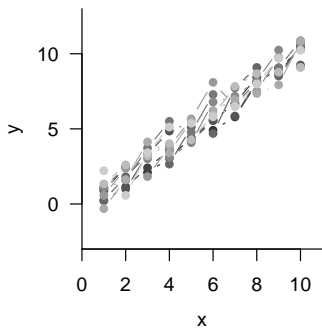
```
# Multivariate Normalverteilung und Matrizenrechnung
library(MASS)
library(matlib)

# Modellformulierung
n      = 10                                # Anzahl von Datenpunkten
p      = 2                                 # Anzahl von Betparametern
x      = 1:n                               # Prädiktorwerte
X      = matrix(c(rep(1,n),x), nrow = n)   # Designmatrix
I_n    = diag(n)                          # n x n Einheitsmatrix
beta   = matrix(c(0,1), nrow = p)         # wahrer, aber unbekannter, Betaparameter
sigsqr = .5                               # wahrer, aber unbekannter, Varianzparameter

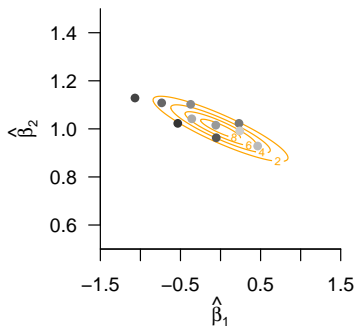
# Frequentistische Simulation
nsim   = 10                                # Anzahl Realisierungen n-dimensionaler ZV
y      = matrix(rep(NaN,n*nsim), nrow = n) # y Realisierungsarray
beta_hat = matrix(rep(NaN,p*nsim), nrow = p) # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y[,i]      = mvrnorm(1, X %>% beta, sigsqr*I_n) # eine Realisierung n-dimensionaler ZV
  beta_hat[,i] = inv(t(X) %>% X) %>% t(X) %>% y[,i] # \hat{\beta} = (X^T X)^{-1} X^T y
}
```

Beispiel (2) Einfache lineare Regression

$$y \sim (X\beta, \sigma^2 I_n)$$



$$\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1})$$



Theorem (Frequentistische Verteilung des Varianzparameterschätzers)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (16)$$

das ALM in generative Form. Weiterhin sei

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (17)$$

der Varianzparameterschätzer. Dann gilt

$$\frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n - p) \quad (18)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Da es sich bei $(y - X\hat{\beta})^T (y - X\hat{\beta})$ um eine Summe normalverteilter Zufallsvariablen handelt, liegt die χ^2 -Verteilung im Lichte der χ^2 Transformation aus Einheit (7) Transformationen der Normalverteilung in Wahrscheinlichkeitstheorie und Frequentistische Inferenz nahe.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \quad (19)$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen. Wir haben bereits gesehen, dass in diesem Fall $\hat{\beta}$ mit dem Stichprobenmittel \bar{y} identisch ist und dass $\hat{\sigma}^2$ mit der Stichprobenvarianz s_y^2 übereinstimmt.

In Einheit (11) Konfidenzintervalle von Wahrscheinlichkeitstheorie und Frequentistische Inferenz hatten wir für den Fall von n unabhängig und identisch normalverteilten Zufallsvariablen die Statistik

$$U := \frac{n-1}{\sigma^2} s^2 \quad (20)$$

definiert und festgehalten, dass

$$U \sim \chi^2(n-1). \quad (21)$$

Offenbar ist U für $p = 1$ mit der im obigen Theorem betrachteten Zufallsvariable $\frac{n-p}{\sigma^2} \hat{\sigma}^2$ identisch.

Schätzerverteilungen

Beispiel (2) Einfache lineare Regression

```
# Libraries
library(MASS)
library(matlib)

# Modellformulierung
n      = 10
p      = 2
x      = 1:n
X      = matrix(c(rep(1,n),x), nrow = n)
I_n    = diag(n)
beta   = matrix(c(0,1), nrow = p)
sigsqr = .5

# multivariate Normalverteilung
# Matrizenrechnung

# Anzahl von Datenpunkten
# Anzahl von Betparametern
# Prädiktorwerte
# Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter, Betaparameter
# wahrer, aber unbekannter, Varianzparameter

# Frequentistische Simulation
nsim   = 1e3
y      = matrix(rep(NaN,n*nsim), nrow = n)
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)
sigsqr_hat = rep(NaN, nsim)
for(i in 1:nsim){
  y[,i]      = mvrnorm(1, X %*% beta, sigsqr*I_n)
  beta_hat[,i] = inv(t(X) %*% X) %*% t(X) %*% y[,i]
  eps_hat    = y[,i] - X %*% beta_hat[,i]
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat)/(n-p)
}
U = ((n-p)/sigsqr)*sigsqr_hat

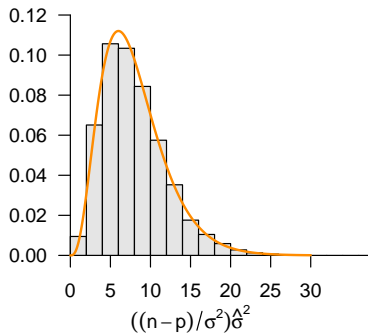
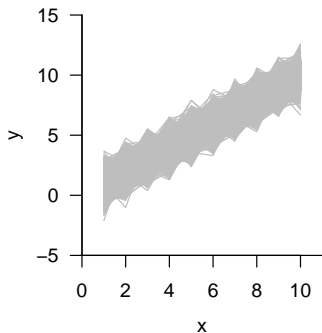
# Anzahl Realisierungen n-dimensionaler ZV
# y Realisierungsarray
# \hat{\beta} Realisierungsarray
# \hat{\sigma}^2 Realisierungsarray
# eine Realisierung n-dimensionaler ZV
# \hat{\beta} = (X^T X)^{-1} X^T y
# \hat{\epsilon} = y - X \hat{\beta}
# \hat{\sigma}^2 = \hat{\epsilon}^T \hat{\epsilon} / (n-p)

# \chi^2 verteilte Zufallsvariable
```

Beispiel (2) Einfache lineare Regression

$$y \sim (X\beta, \sigma^2 I_n)$$

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p)$$



Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Überblick

- T-Statistiken quantifizieren die geschätzten Effekte in $\hat{\beta}$ in bezug auf die Residualvariabilität $\hat{\sigma}^2$. T-Statistiken sind also als Signal-zu-Rauschen Verhältnisse (signal-to-noise ratios) zu verstehen.
- T-Statistiken erlauben die Frequentistische Evaluation von linearen Kombinationen der $\hat{\beta}$ Komponenten. Eine einfache Linearkombination von $\hat{\beta}$ ist das Skalarprodukt meinem Einheitsvektor. Solch eine Linearkombination "extrahiert" eine Komponente von $\hat{\beta}$ und erlaubt damit ihre Frequentistische Evaluation.
- T-Teststatistiken erlauben weiterhin die Evaluation von Linearkombinationen von $\hat{\beta}$ im Sinne Frequentistischer Hypothesentests. Wir betrachten hier zunächst die nur die funktionale Form von T-Teststatistiken und ihre Frequentistische Verteilung. Der Einsatz von T-Teststatistiken im Rahmen von Einstich- und Zweistichproben T-Tests ist Inhalt von (9) T-Tests.

Definition (t -Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

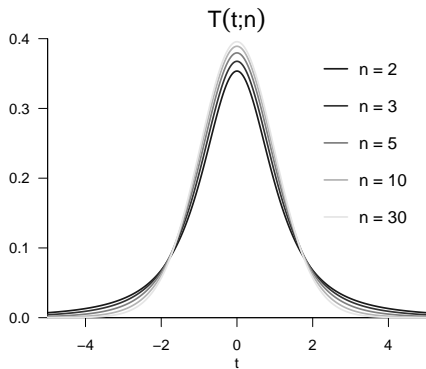
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad (22)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass T einer t -Verteilung mit n Freiheitsgraden unterliegt und nennen T eine t -Zufallsvariable mit n Freiheitsgraden. Wir kürzen dies mit $T \sim t(n)$ ab. Die WDF einer t -Zufallsvariable bezeichnen wir mit

$$T(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \quad (23)$$

Bemerkungen

- Die Verteilung ist um 0 symmetrisch
- Steigendes n verschiebt Wahrscheinlichkeitsmasse aus den Ausläufen zum Zentrum.
- Ab $n = 30$ gilt $T(t; n) \approx N(0, 1)$.



Theorem (T-Transformation)

$Z \sim N(0, 1)$ sei eine Z -Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 -Zufallsvariable mit n Freiheitsgraden, und Z und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{Z}{\sqrt{U/n}} \quad (24)$$

eine t -verteilte Zufallsvariable mit n Freiheitsgraden, es gilt also $T \sim t(n)$.

Bemerkungen

- Das Theorem geht auf Student (1908) zurück.
- Das Theorem ist das zentrale Resultat der Frequentistischen Statistik.
- Zabell (2008) gibt hierzu einen historischen Überblick.
- Die T-Statistik und T-Tests sind wichtige Anwendungsfälle.

Definition (Nichtzentrale t -Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

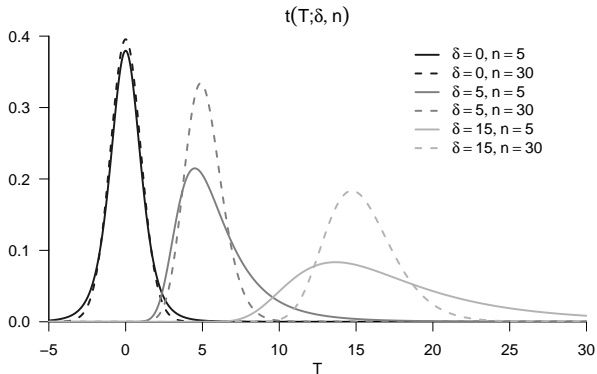
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) (n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \quad (25)$$

Dann sagen wir, dass T einer nichtzentralen t -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n unterliegt und nennen T eine *nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n* . Wir kürzen dies mit $t(\delta, n)$ ab. Die WDF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $t(T; \delta, n)$. Die KVF und inverse KVF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $\psi(\cdot; \delta, n)$ und $\psi^{-1}(\cdot; \delta, n)$, respektive.

Bemerkungen

- Eine nichtzentrale t -Zufallsvariable mit $\delta = 0$ ist eine t -Zufallsvariable.
- Es gilt also $t(T; 0, n) = t(T; n)$.
- Die funktionale Form der WDF findet sich zum Beispiel in Lehmann (1986), Seite 254, Gl. (80).

Wahrscheinlichkeitsdichtefunktionen nichtzentraler t -Verteilungen



Theorem (Nichtzentrale T-Transformation)

$X \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n , und X und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{X}{\sqrt{U/n}} \quad (26)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n , es gilt also $T \sim t(\mu, n)$

Bemerkung

- Wir verzichten auf einen Beweis.

Definition (T-Statistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (27)$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (28)$$

die Betaparameter- und Varianzparameterschätzer, respektive. Dann ist für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ die *T-Statistik* definiert als

$$T := \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \quad (29)$$

Bemerkungen

- Die T-Statistik hängt via $\hat{\beta}$ und $\hat{\sigma}^2$ von den Daten y ab.
- Der Kontrastgewichtsvektor projiziert $\hat{\beta}$ auf einen Skalar $c^T \hat{\beta}$.
- Die folgende Intuition zur T-Statistik ist im Allgemeinen hilfreich

$$T = \frac{\text{Geschätzte Effektstärke}}{\text{Geschätzte stichprobenumfangskalierte Datenvariabilität}} \quad (30)$$

- Intuitiv repräsentiert eine T-Statistik also ein Signal-zu-Rauschen Verhältnis.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \quad (31)$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen und es sei $c := 1$. Dann gilt für die T-Statistik

$$T = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{\mathbf{1}^T \bar{y}}{\sqrt{s_y^2 \mathbf{1}^T (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}}} = \sqrt{n} \frac{\bar{y}}{s_y} \quad (32)$$

was der Einstichproben-T-Teststatistik für den Fall $\mu_0 = 0$ entspricht (vgl. Einheit (13) Einstichproben-T-Tests in Wahrscheinlichkeitstheorie und Frequentistische Inferenz und Einheit (9) T-Tests in Allgemeines Lineares Modell). Die hier betrachtete T-Statistik nimmt hohe Werte für hohe Werte von \bar{y} (Effekt), kleine Werte von s_y^2 (Datenvariabilität) und hohe Werte von n (Stichprobenumfang) an.

Eine beliebige Definition in diesem Zusammenhang ist *Cohen's d* als *Effektstärkenmaß*. Es gilt

$$d := \frac{\bar{y}}{s_y}, \quad (33)$$

so dass

$$T = \sqrt{n}d \text{ bzw. } d = \frac{1}{\sqrt{n}} T. \quad (34)$$

Cohen's d ist also ein stichprobenumfangunabhängiges Signal-zu-Rauschen Verhältnis.

Theorem (T-Teststatistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (35)$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (36)$$

die Betaparameter- und Varianzparameterschätzer, respektive. Schließlich sei für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ und einen *Nullhypothesenbetaparameter* $\beta_0 \in \mathbb{R}^p$ die *T-Teststatistik* definiert als

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \quad (37)$$

Dann gilt

$$T \sim t(\delta, n - p) \text{ mit } \delta := \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} \quad (38)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- T ist eine Funktion der Parameterschätzer, δ ist eine Funktion der wahren, aber unbekanntem, Parameter
- Für $c^T \beta = c^T \beta_0$, also bei Zutreffen der Nullhypothese, gilt $\delta = 0$ und damit $T \sim t(n - p)$.
- Für $c^T \beta \neq c^T \beta_0$ kann die Verteilung von T zur Herleitung von Powerfunktionen benutzt werden..

T-Statistiken

Simulation (1) Unabhängig und identische normalverteilte Zufallsvariablen

Wahre, aber unbekannte, Hypothesenszenarien $c^T \beta = c^T \beta_0$ und $c^T \beta \neq c^T \beta_0$

```
# Libraries
library(MASS) # multivariate Normalverteilung

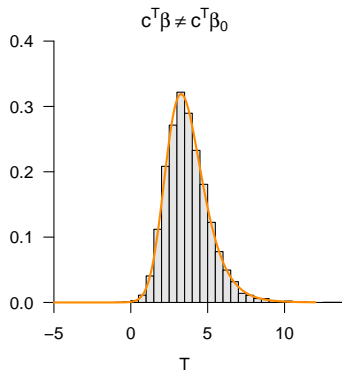
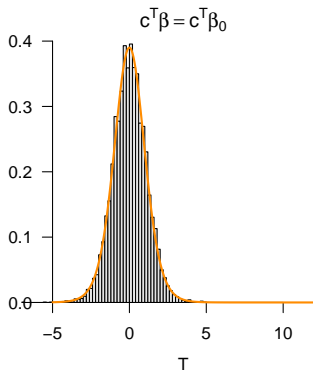
# Modellformulierung
n = 12 # Anzahl von Datenpunkten
p = 1 # Anzahl von Betaparametern
X = matrix(c(rep(1,n)), nrow = n) # Designmatrix
I_n = diag(n) # Einheitsmatrix
beta = c(0,1) # wahre , aber unbekannte , Betaparameter
nscn = length(beta) # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter
c = 1 # Kontrastvektor von Interesse
beta_0 = 0 # Nullhypothesenbetaparameter

# Frequentistische Simulation
nsim = 1e4 # Anzahl Simulationen
delta = rep(NA, nscn) # Anzahl Nichtzentralitätsparameter
Tee = matrix(rep(NA, nscn*nsim), ncol = nscn) # T-Teststatistik Realisierungsarray
for(s in 1:nscn){ # Hypothesenszenarien
  delta[s] = ((t(c) %*% beta[s] - t(c) %*% beta_0)/ # Nichtzentralitätsparameter
             sqrt(sigsqr*t(c)%*%solve(t(X)%*%X)%*%c))
  for(i in 1:nsim){ # Simulationsiterationen
    y = mvrnorm(1, X %*% beta[s], sigsqr*I_n) # y
    beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # \hat{\beta}
    eps_hat = y - X %*% beta_hat # \hat{\epsilon}
    sigsqr_hat = (t(eps_hat) %*% eps_hat)/(n-p) # \hat{\sigma}^2
    Tee[i,s] = ((t(c) %*% beta_hat - t(c) %*% beta_0)/ # T
               sqrt(sigsqr_hat*t(c)%*%solve(t(X)%*%X)%*%c))
  }
}
```

T-Statistiken

Simulation (1) Unabhängig und identische normalverteilte Zufallsvariablen

Wahre, aber unbekannte, Hypothesenszenarien $c^T \beta = c^T \beta_0$ und $c^T \beta \neq c^T \beta_0$



T-Statistiken

Simulation (1) Einfache lineare Regression

Wahre, aber unbekannte, Hypothesenszenarien $c^T \beta = c^T \beta_0$ und $c^T \beta \neq c^T \beta_0$

```
# Libraries
library(MASS) # multivariate Normalverteilung
library(matlib) # Matrizenrechnung

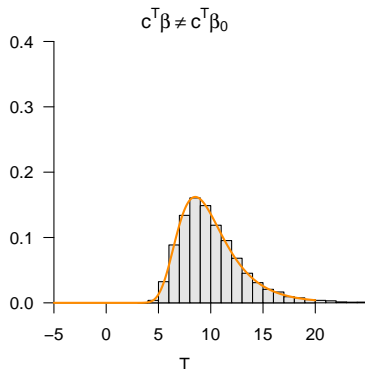
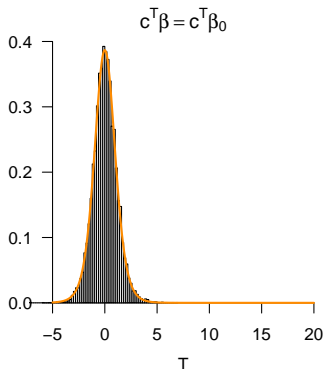
# Modellformulierung
n = 10 # Anzahl von Datenpunkten
p = 2 # Anzahl von Betaparametern
x = 1:n # Prädiktorwerte
X = matrix(c(rep(1,n),x), ncol = p) # Designmatrix
I_n = diag(n) # Einheitsmatrix
beta = matrix(c(1,0,1,1), nrow = 2) # wahre , aber unbekannte , Betaparameter
nscn = ncol(beta) # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter
c = matrix(c(0,1), nrow = 2) # Kontrastvektor von Interesse
beta_0 = matrix(c(0,0), nrow = 2) # Nullhypothesenbetaparameter

# Frequentistische Simulation
nsim = 1e4 # Anzahl Simulationen
delta = rep(NaN, nscn) # Anzahl Nichtzentralitätsparameter
Tee = matrix(rep(NaN, nscn*nsim), ncol = nscn) # T-Teststatistik Realisierungsarray
for(s in 1:nscn){ # Hypothesenszenarien
  delta[s] = ((t(c) %*% beta[,s] - t(c) %*% beta_0)/
    sqrt(sigsqr*t(c)%*%inv(t(X)%*%X)%*%c)) # Nichtzentralitätsparameter
  for(i in 1:nsim){ # Simulationsiterationen
    y = mvrnorm(1, X %*% beta[,s], sigsqr*I_n) # y
    beta_hat = inv(t(X) %*% X) %*% t(X) %*% y # \hat{\beta}
    eps_hat = y - X %*% beta_hat # \hat{\epsilon}
    sigsqr_hat = (t(eps_hat) %*% eps_hat)/(n-p) # \hat{\sigma}^2
    Tee[i,s] = ((t(c) %*% beta_hat - t(c) %*% beta_0)/ # T
      sqrt(sigsqr_hat*t(c)%*%inv(t(X)%*%X)%*%c))
  }
}
```


T-Statistiken

Simulation (1) Einfache lineare Regression

Wahre, aber unbekannte, Hypothesenszenarien $c^T \beta = c^T \beta_0$ und $c^T \beta \neq c^T \beta_0$



Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Überblick

- F-Statistiken können als Modellvergleichskriterien zwischen ALMs verstanden werden. Im Gegensatz zu T-Statistiken kann das Ziel der Berechnung von F-Statistiken sein, nicht nur Linearkombinationen von Betaparameterschätzwerten probabilistisch zu evaluieren, sondern die Modellanpassung an einen Datensatz insgesamt zu evaluieren.
- Wir motivieren die funktionale Form der F-Statistik als Residualquadratsummendifferenz hier vor dem Hintergrund eine Likelihood-Quotienten zum Modellvergleich. Die Modellvergleichskapazität von F-Statistiken ist allerdings etwas beschränkt, da sich die F-Statistik auf ALMs und insbesondere geschachtelte ALMs bezieht, in denen ein Modell Bestandteil eines anderen Modells ist.
- F-Statistiken bilden die Grundlage für Hypothesentests im Rahmen der Varianzanalyse (Einheiten (10) Einfaktorielle Varianzanalyse, (11) Zweifaktorielle Varianzanalyse, (13) Kovarianzanalyse). Der Einsatz von F-Statistiken ist aber *perse* nicht auf Varianzanalysen beschränkt.

Definition (f -Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

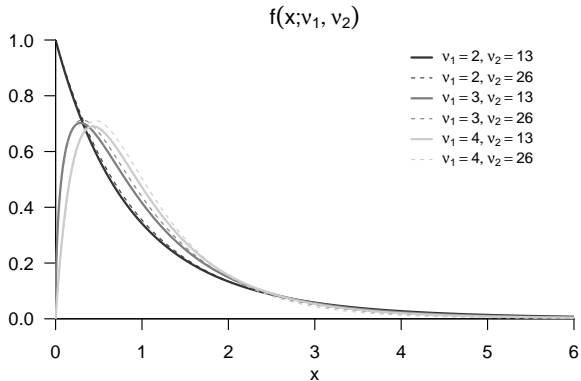
$$p_X : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p_X(x) := \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2} - 1}}{(\nu_1 x + \nu_2)^{\frac{\nu_1 + \nu_2}{2}}}, \quad (39)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer f -Verteilung mit Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen X eine f -Zufallsvariable mit Freiheitsgradparametern ν_1 und ν_2 . Wir kürzen dies mit $X \sim f(\nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f -Zufallsvariable bezeichnen wir mit $f(x; \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer f -Zufallsvariable bezeichnen wir mit $\varphi(x; \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer f -Zufallsvariable bezeichnen wir mit $\varphi^{-1}(x; \nu_1, \nu_2)$.

Bemerkungen

- f -Zufallsvariablen sind nach Ronald A. Fisher benannt.
- George W. Snedecor hat die KVF der F -Verteilung wohl 1934 basierend auf Arbeiten von Fisher tabuliert.

Wahrscheinlichkeitsdichtefunktionen von f -Verteilungen



Theorem (F -Transformation)

$V \sim \chi^2(\nu_1)$ und $W \sim \chi^2(\nu_2)$ seien zwei unabhängige χ^2 -Zufallsvariablen mit ν_1 und ν_2 Freiheitsgraden, respektive. Dann ist die Zufallsvariable

$$F := \frac{V/\nu_1}{W/\nu_2} \quad (40)$$

eine f -verteilte Zufallsvariable mit ν_1, ν_2 Freiheitsgraden, es gilt also $F \sim f(\nu_1, \nu_2)$.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Das Theorem kann bewiesen werden, in dem man zunächst ein Transformationstheorem für Quotienten von Zufallsvariablen mithilfe des multivariaten Transformationstheorems und Marginalisierung herleitet und dieses Theorem dann auf die WDF von χ^2 -verteilten ZVen anwendet. Dabei ist die Regel zur Integration durch Substitution von zentraler Bedeutung.

Definition (Nichtzentrale f -Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

$$p_X : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto$$

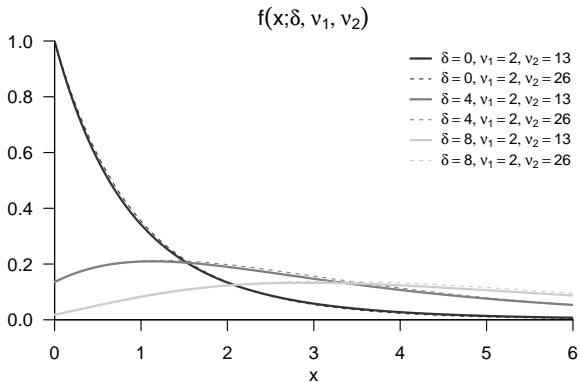
$$p_X(x) := \sum_{k=0}^{\infty} \frac{e^{-\delta/2} (\delta/2)^k}{\frac{\Gamma(\nu_2/2)\Gamma(\nu_1/2+k)}{\Gamma(\nu_2/2+\nu_1/2+k)} k!} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2+k} \left(\frac{\nu_2}{\nu_2 + \nu_1 x}\right)^{(\nu_1+\nu_2)/2+k} x^{\nu_1/2-1+k} \quad (41)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer nichtzentralen f -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen X eine nichtzentrale f -Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 . Wir kürzen dies mit $X \sim f(\delta, \nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f -Zufallsvariable bezeichnen wir mit $f(x; \delta, \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer nichtzentralen f -Zufallsvariable bezeichnen wir mit $\varphi(x; \delta, \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer nichtzentralen f -Zufallsvariable bezeichnen wir mit $\varphi^{-1}(x; \delta, \nu_1, \nu_2)$.

Bemerkungen

- Es gilt $f(0, \nu_1, \nu_2) = f(\nu_1, \nu_2)$.

WDFen von nichtzentralen f -Verteilungen



Theorem (Nichtzentrale F -Transformation)

$V \sim \chi^2(\delta, \nu_1)$ und $W \sim \chi^2(\nu_2)$ seien eine nichtzentrale χ^2 Zufallsvariable und eine χ^2 -Zufallsvariablen mit ν_1 und ν_2 Freiheitsgraden, respektive, und V und W seien unabhängig. Dann ist die Zufallsvariable

$$F := \frac{V/\nu_1}{W/\nu_2} \quad (42)$$

eine nichtzentral f -verteilte Zufallsvariable mit Nichtzentralitätsparameter δ und ν_1, ν_2 Freiheitsgraden, es gilt also $F \sim f(\delta, \nu_1, \nu_2)$.

Bemerkungen

- Wir verzichten auf einen Beweis.

Definition (Vollständiges und reduziertes Modell)

Für $p > 1$ mit $p = p_1 + p_2$ seien

$$X := \begin{pmatrix} X_1 & X_2 \end{pmatrix} \in \mathbb{R}^{n \times p} \text{ mit } X_1 \in \mathbb{R}^{n \times p_1} \text{ und } X_2 \in \mathbb{R}^{n \times p_2}, \quad (43)$$

sowie

$$\beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^p \text{ mit } \beta_1 \in \mathbb{R}^{p_1} \text{ und } \beta_2 \in \mathbb{R}^{p_2} \quad (44)$$

Partitionierungen einer $n \times p$ Designmatrix und eines p -dimensionalen Betaparametervektors. Dann nennen wir

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (45)$$

das *vollständige Modell* und

$$y = X_1\beta_1 + \varepsilon_1 \text{ mit } \varepsilon_1 \sim N(0_n, \sigma^2 I_n) \quad (46)$$

das *reduzierte Modell* und sprechen von einer *Partitionierung eines (vollständigen) Modells*.

Bemerkungen

- Man sagt auch, dass das reduzierte Modell im vollständigen Modell *geschachtelt (nested)* ist.

Motivation der F-Statistik | Modellpartitionierung und Likelihood Ratios

Definition (Likelihood-Quotient von vollständigem und reduzierten Modell)

Für $p = p_1 + p_2$, $p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und $\sigma^2 > 0$ als bekannt vorausgesetzt. Weiterhin seien

$$L : \mathbb{R}^p \rightarrow \mathbb{R}_{>0}, \beta \mapsto L(\beta) := p_\beta(v) = N(v; X\beta, \sigma^2 I_n) \quad (47)$$

und

$$L_1 : \mathbb{R}^{p_1} \rightarrow \mathbb{R}_{>0}, \beta_1 \mapsto L_1(\beta_1) := p_{\beta_1}(v) = N(v; X_1\beta_1, \sigma^2 I_n) \quad (48)$$

die Likelihood-Funktionen von vollständigem und reduziertem Modell, respektive. Dann ist für einen Datenvektor $v \in \mathbb{R}^n$ der *Likelihood-Quotient von vollständigem und reduziertem Modell* gegeben als

$$:= \frac{\max_{\beta \in \mathbb{R}^p} p_\beta(v)}{\max_{\beta_1 \in \mathbb{R}^{p_1}} p_{\beta_1}(v)} = \frac{\max_{\beta \in \mathbb{R}^p} L(\beta)}{\max_{\beta_1 \in \mathbb{R}^{p_1}} L_1(\beta_1)} = \frac{L(\hat{\beta})}{L_1(\hat{\beta}_1)} \quad (49)$$

Motivation der F-Statistik | Modellpartitionierung und Likelihood Ratios

Bemerkungen

- Der Likelihood-Quotient setzt die Wahrscheinlichkeitsdichten eines beobachteten Datensatzes $v \in \mathbb{R}^n$ unter dem vollständigem und dem reduzierten Modell nach Optimierung der jeweiligen Modellbetaparameter ins Verhältnis.
- Ein hoher Wert des Likelihood-Quotienten entspricht einer höheren Wahrscheinlichkeit(sdichte) des beobachteten Datensatzes $v \in \mathbb{R}^n$ unter dem optimierten vollständigem Modell und vice versa.
- Die Wahrscheinlichkeiten beobachtbarer Daten nach Modellschätzung unter verschiedenen Modellen zu betrachten ist ein allgemeines Vorgehen zum Modellvergleich. Letztlich erlaubt dieses Vorgehen, verschiedene wissenschaftliche Theorien (= Modelle) über die Genese beobachtbarer Daten quantitativ zu vergleichen und die damit verbundene Unsicherheit zu quantifizieren.
- Modellvergleiche sind ein zentrales Thema in der Bayesianischen Statistik, die die Logik von Likelihood-Quotienten zum Beispiel unter den Begriffen der Bayes Factors oder der des Bayesian Information Criteria auf arbiträre probabilistische Modelle, d.h. nicht nur geschachtelte ALMs generalisiert.
- Allerdings sind, wie hier gesehen, Modellvergleiche auch im Rahmen der Frequentistischen Statistik möglich und sinnvoll, Modellvergleiche sind also kein Alleinstellungsmerkmal der Bayesianischen gegenüber der Frequentistischen Statistik.

Theorem (Likelihood-Quotient und Residualquadratsummendifferenz)

Für $p = p_1 + p_2$, $p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und $\sigma^2 > 0$ als bekannt vorausgesetzt. Weiterhin seien

$$\hat{\epsilon} := y - X\hat{\beta} \text{ und } \hat{\epsilon}_1 := y - X_1\hat{\beta}_1 \quad (50)$$

die Residuenvektoren des vollständigen und des reduzierten Modells, respektive. Dann gilt

$$\ln = \frac{\hat{\epsilon}_1^T \hat{\epsilon}_1 - \hat{\epsilon}^T \hat{\epsilon}}{2\sigma^2} \quad (51)$$

Bemerkungen

- \ln und die F-Statistik sind nicht identisch, aber eng verwandt.
- $\hat{\epsilon}_1^T \hat{\epsilon}_1$ ist die Residualquadratsumme (RQS), die bei Schätzung des reduzierten Modells von y resultiert.
- $\hat{\epsilon}^T \hat{\epsilon}$ ist die RQS, die bei Schätzung des vollständigen Modells anhand desselben y resultiert.
- Für $\beta_2 := 0_{p_2}$ sind vollständiges und reduziertes Modell identisch, es muss also $\hat{\epsilon}_1^T \hat{\epsilon}_1 \geq \hat{\epsilon}^T \hat{\epsilon}$ gelten.
- $\hat{\epsilon}_1^T \hat{\epsilon}_1 - \hat{\epsilon}^T \hat{\epsilon}$ misst, inwieweit für $\beta_2 \neq 0_{p_2}$ die Spalten von X_2 die RQS verringern.
- Ist $\hat{\epsilon}_1^T \hat{\epsilon}_1$ groß, aber $\hat{\epsilon}^T \hat{\epsilon}$ klein, so ist \ln groß und der Beitrag der X_2 -Spalten zur Datenerklärung groß.
- Sind $\hat{\epsilon}_1^T \hat{\epsilon}_1$ und $\hat{\epsilon}^T \hat{\epsilon}$ gleich groß, also \ln klein, ist und der Beitrag der X_2 -Spalten zur Datenerklärung klein.
- "Groß" und "klein" von $\hat{\epsilon}_1^T \hat{\epsilon}_1$ und $\hat{\epsilon}^T \hat{\epsilon}$ werden in Relation zur doppelten Fehlervektorvarianz σ^2 gemessen.

Definition (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (52)$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \quad (53)$$

mit $p = p_1 + p_2$ gegeben. Weiterhin seien mit

$$\hat{\beta}_1 := (X_1^T X_1)^{-1} X_1^T y \text{ und } \hat{\beta} := (X^T X)^{-1} X^T y \quad (54)$$

die Residuenvektoren

$$\hat{\varepsilon}_1 := y - X_1 \hat{\beta}_1 \text{ und } \hat{\varepsilon} := y - X \hat{\beta} \quad (55)$$

definiert. Dann ist die F-Statistik definiert als

$$F := \frac{\frac{\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_2}}{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p}} \quad (56)$$

Bemerkungen

- Der Zähler der F-Statistik

$$\frac{\hat{\epsilon}_1^T \hat{\epsilon}_1 - \hat{\epsilon}^T \hat{\epsilon}}{p_2} \quad (57)$$

misst, inwieweit die p_2 Regressoren in X_2 die Residualquadratsumme reduzieren und zwar im Verhältnis zur Anzahl dieser Regressoren. Das heißt, dass bei gleicher Größe der Residualquadratsummenreduktion (und gleichem Nenner) ein größerer F Wert resultiert, wenn diese durch weniger zusätzliche Regressoren resultiert, also p_2 klein ist (und vice versa). Im Sinne der Anzahl der Spalten von X und der entsprechenden Komponenten von β favorisiert die F -Statistik also weniger "komplexe" Modelle.

- Für den Nenner der F-Statistik gilt

$$\frac{\hat{\epsilon}^T \hat{\epsilon}}{n - p} = \hat{\sigma}^2, \quad (58)$$

wobei $\hat{\sigma}^2$ hier der aufgrund des vollständigen Modells geschätzte Schätzer von σ^2 ist. Werden die Daten tatsächlich unter dem reduzierten Modell generiert, so kann das vollständige Modell dies durch $\hat{\beta}_2 \approx 0_{p_2}$ abbilden und erreicht eine ähnliche σ^2 Schätzung wie das reduzierte Modell. Werden die Daten de-facto unter dem vollständigem Modell generiert, so ist $\hat{\epsilon}^T \hat{\epsilon} / (n - p)$ ein besserer Schätzer von σ^2 als $\hat{\epsilon}_1^T \hat{\epsilon}_1 / (n - p)$, da sich für diesen Datenvariabilität, die nicht durch die p_1 Regressoren in X_1 erklärt wird, in der Schätzung von σ^2 widerspiegeln würde. Der Nenner der F-Statistik ist also in beiden Fällen der sinnvollere Schätzer von σ^2 .

- Zusammengefasst misst die F-Statistik also die Residualquadratsummenreduktion durch die p_2 Regressoren in X_2 gegenüber den p_1 Regressoren in X_1 pro Datenvariabilitäts (σ^2)- und Regressor (p_2)-Einheit.

F-Statistiken

Beispiel (1) Einfache lineare Regression

$$X = (X_1 \quad X_2), X_1 := 1_n, X_2 := (x_1, \dots, x_n)^T, \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 = \text{Interzept}, \beta_2 = \text{Steigung} \quad (59)$$

```
# Libraries
library(MASS) # Multivariate Normalverteilung

# Modellformulierung
nmod = 2 # Anzahl Modelle
n = 10 # Anzahl Datenpunkte
p = 2 # Anzahl Betaparameter
p_1 = 1 # Anzahl Betaparameter reduziertes Modell
p_2 = 1 # Anzahl zusätzlicher Betaparameter vollständiges Modell
p = p_1 + p_2 # Anzahl Betaparameter im vollständigem Modell
x = 1:n # Prädiktorwerte
X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix des vollständigen Modells
X_1 = X[,1] # Designmatrix des reduzierten Modells
I_n = diag(n) # n x n Einheitsmatrix
beta = matrix(c(1,0,1,.5), nrow = 2) # wahre, aber unbekannte, Betaparameter
nscn = ncol(beta) # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter

# Modellsimulation und Evaluierung
Eff = matrix(rep(NA,n), nrow = nscn) # F-Statistik Realisierungsarray
for(s in 1:nscn){ # Szenarieniterationen
  y = mvrnorm(1, X %*%beta[,s], sigsqr*I_n) # Datenrealisierung
  beta_hat_1 = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y # Betaparameterschätzer reduziertes Modell
  beta_hat = solve(t(X) %*%X )%*%t(X) %*%y # Betaparameterschätzer vollständiges Modell
  eps_1_hat = y-X_1%*%beta_hat_1 # Residuenvektor reduziertes Modell
  eps_hat = y-X%*%beta_hat # Residuenvektor vollständiges Modell
  eps_1_eps_1_hat = t(eps_1_hat) %*% eps_1_hat # RQS reduziertes Modell
  eps_eps_hat = t(eps_hat) %*% eps_hat # RQS vollständiges Modell
  Eff[s] = (((eps_1_eps_1_hat-eps_eps_hat)/p_2)/ # F-Statistik
            (eps_eps_hat/(n-p)))}
```

> F-Statistik für beta_2 = 0_{p_2}: 0.404

> F-Statistik für beta_2 != 0_{p_2}: 24.4

Theorem (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (60)$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \quad (61)$$

mit $p = p_1 + p_2$ gegeben. Schließlich sei

$$K := \begin{pmatrix} 0_{p_1} \\ 1_{p_2} \end{pmatrix} \in \mathbb{R}^p \quad (62)$$

ein Kontrastgewichtsvektor. Dann gilt

$$F \sim f(\delta, p_2, n - p) \text{ mit } \delta := \frac{K^T \beta \left(K^T (X^T X)^{-1} K \right)^{-1} K^T \beta}{\sigma^2} \quad (63)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- F ist eine Funktion der Parameterschätzer, δ ist eine Funktion der wahren, aber unbekanntem Parameter
- Diese Verteilung von F kann zum Nullhypothestesten und zur Powerfunktionsevaluation genutzt werden.

F-Statistiken

```
# Libraries
library(MASS)

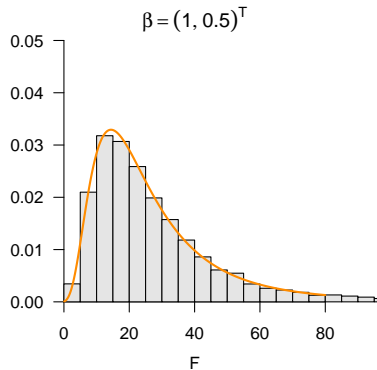
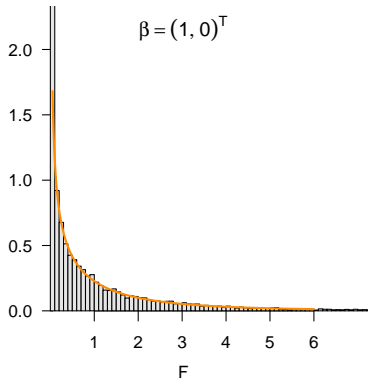
# Modellformulierung
nmod = 2
n = 10
p = 2
p_1 = 1
p_2 = 1
p = p_1 + p_2
x = 1:n
X = matrix(c(rep(1,n),x), nrow = n)
X_1 = X[,1]
I_n = diag(n)
beta = matrix(c(1,0,1,.5), nrow = 2)
nscn = ncol(beta)
sigsqr = 1
K = matrix((c(0,1)), nrow = 2)

# Multivariate Normalverteilung

# Anzahl Modelle
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Anzahl Betaparameter im reduzierten Modell
# Anzahl zusätzlicher Betaparameter im vollständigen Modell
# Anzahl Betaparameter im vollständigem Modell
# Prädiktorwerte
# Designmatrix des vollständigen Modells
# Designmatrix des reduzierten Modells
# n x n Einheitsmatrix
# wahre, aber unbekannte, Betaparameter
# Anzahl wahrer, aber unbekannter, Hypothesenszenarien
# wahrer, aber unbekannter, Varianzparameter

# Frequentistische Simulation
nsim = 1e4
delta = rep(NA,nscn)
Eff = matrix(rep(NA,nscn*nsim), nrow = nscn)
for(s in 1:nsim){
  delta[s] = (t(t(K)%*beta[,s])%*%
             solve(t(K)%*%solve(t(X)%*%X)%*%K) %*%
             (t(K)%*beta[,s])/sigsqr)
  for(i in 1:nsim){
    y = mvrnorm(1, X %*%beta[,s], sigsqr*I_n)
    beta_hat_1 = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y
    beta_hat = solve(t(X) %*%X )%*%t(X) %*%y
    eps_1_hat = y-X_1%*beta_hat_1
    eps_hat = y-X%*beta_hat
    eps_1_eps_1_hat = t(eps_1_hat) %*% eps_1_hat
    eps_eps_hat = t(eps_hat) %*% eps_hat
    Eff[s,i] = (((eps_1_eps_1_hat-eps_eps_hat)/p_2)/
               (eps_eps_hat/(n-p)))
  }
}
```

Beispiel (1) Einfache lineare Regression



Schlussbemerkungen

- Die Theorie von T- und F-Statistiken kann unter dem Gesichtspunkt von linearen Transformationen von Betaparametern verallgemeinert und integriert werden. Dabei betrachtet man Hypothesen der Form

$$K\beta = \beta_0, \tag{64}$$

wobei $K \in \mathbb{R}^{s \times p}$ eine beliebige Matrix ist. Zum Beispiel ergibt sich für $K \in \mathbb{R}^{1 \times p}$ der hier betrachtete Fall von T-Statistiken und für $K = I_p \in \mathbb{R}^{p \times p}$ der hier betrachtete Fall von F-Statistiken.

- Man kann weiterhin zeigen, dass im Wesentlichen gilt

$$F = T^2, \tag{65}$$

dass also auch das Quadrat einer T-Statistik f -verteilt ist und T-Statistiken damit spezielle F-Statistiken sind.

- Dennoch wird in der Anwendung sehr stark zwischen T- und F-Statistiken unterschieden und es ist sinnvoll, sich der unterschiedlichen Anwendungsfälle von T- und F-Statistiken bewusst zu sein. Eine sehr gute Einführung in die allgemeine Kontrasttheorie gibt Searle (1971), Chapter 3.6.

Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers wieder.
2. Geben Sie die Verteilung des Betaparameterschätzers im Szenario von n u.i.v. Zufallsvariablen wieder.
3. Geben Sie das Theorem zur Frequentistischen Verteilung des Varianzparameterschätzers wieder.
4. Geben Sie die Verteilung des skalierten Varianzparameterschätzers bei n u.i.v. Zufallsvariablen wieder.
5. Skizzieren Sie die WDFen von t -Zufallsvariablen mit 2, 10 und 30 Freiheitsgraden.
6. Skizzieren Sie die WDFen von nichtzentralen t -Zufallsvariablen mit Nichtzentralitätsparametern 0,5 und 15.
7. Geben Sie die Definition der T-Statistik wieder.
8. Erläutern Sie die Definition der T-Statistik.
9. Warum kann die T-Statistik als Signal-zu-Rauschen Verhältnis interpretiert werden.
10. Geben Sie die Form der T-Statistik im Szenario von n u.i.v. Zufallsvariablen wieder.
11. Erläutern Sie den Zusammenhang der T-Statistik und Cohen's d .
12. Geben Sie die Definition der T-Teststatistik wieder.

Selbstkontrollfragen

13. Erläutern Sie die Definition der T-Teststatistik.
14. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
15. Geben Sie die Definition eines vollständigen und eines reduzierten ALMs wieder.
16. Geben Sie die Definition des Likelihood-Quotienten eines vollständigen und eines reduzierten ALMs wieder.
17. Geben Sie das Theorem zu Likelihood-Quotient und Residualquadratsummendifferenz wieder.
18. Definieren Sie die F-Statistik.
19. Erläutern Sie den Zusammenhang zwischen F-Statistik und Likelihood-Quotient.
20. Definieren Sie und erläutern Sie den Zähler der F-Statistik
21. Definieren Sie und erläutern Sie den Nenner der F-Statistik.
22. Erläutern Sie die F-Statistik.
23. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

References

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Statistics.

Searle, S. R. 1971. *Linear Models*. A Wiley Publication in Mathematical Statistics. New York: Wiley.

Student. 1908. "The Probable Error of a Mean." *Biometrika* 6 (1): 1–25.

Zabell, S. L. 2008. "On Student's 1908 Article 'The Probable Error of a Mean'." *Journal of the American Statistical Association* 103 (481): 1–7. <https://doi.org/10.1198/016214508000000030>.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(8) Studiendesign

Überblick

Datum	Einheit	Thema
08.04.2022	Grundlagen Osterpause	(1) Regression
22.04.2022	Grundlagen	(2) Korrelation
29.04.2022	Grundlagen	(3) Matrizen
06.05.2022	Grundlagen	(4) Normalverteilungen
13.05.2022	Theorie	(5) Modellformulierung
20.05.2022	Theorie	(6) Modellschätzung
27.05.2022	Theorie	(7) Modellevaluation
03.06.2021	Anwendung	(8) Studiendesign
10.06.2021	Anwendung	(9) T-Tests
17.06.2021	Anwendung	(10) Einfaktorielle Varianzanalyse
24.06.2022	Anwendung	(11) Zweifaktorielle Varianzanalyse
01.07.2022	Anwendung	(12) Multiple Regression
08.07.2022	Anwendung	(13) Kovarianzanalyse
14.07.2022	Klausurtermin	
März 2023	Klausurwiederholungstermin	

Studiendesign

- Einführung in die grundlegende Nomenklatur des Studiendesigns
- Grundlegende Logik experimenteller Kontrolle durch Subtraktion
- Grundlegende Logik faktorieller Studiendesigns
- Anwendungsbeispiel für ALM Designs im weiteren Kursverlauf

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Empirische Studie

Eine **empirische Studie** ist gekennzeichnet durch systematische Datenerhebung und/oder Datenanalyse und dient der vorläufigen Beantwortung inhaltlicher Forschungsfragen. Im Rahmen quantitativer empirischer Studien werden Aspekte der Wirklichkeit variiert und gemessen und so als Werte von **Variablen** repräsentiert.

Weitere wichtige und wertvolle Studientypen sind zum Beispiel **theoretische Studien**, die der Weiterentwicklung von wissenschaftlichen Theorien und Modellen dienen, und **Methodenstudien**, die der Weiterentwicklung der wissenschaftlichen Methoden dienen.

Konzeptuelle Variablentypen und Experimentelle Einheiten

Unabhängige Variable (UV)

Etwas, das in einer Studie variiert wird, um seine Auswirkung auf abhängige Variablen zu studieren.

Abhängige Variable (AV)

Etwas, das in einer Studie erfasst wird, um die Auswirkungen der unabhängigen Variablen zu studieren.

Experimentelle Einheit (EE)

Etwas, das der UV ausgesetzt wird und an dem die AV bestimmt wird.

Beispiele

- Einfluss von Psychotherapie Setting (UV) auf Symptomreduktion (AV) bei Patient:innen (EE)
- Einfluss von COVID-19 Impfstofftypen (UV) auf Antikörperlevel (AV) bei Mäusen (EE)
- Einfluss von Düngemitteln (UV) auf Getreidewachstum (AV) auf Versuchsäckern (EE) (Rothamsted Research)

Numerische Variablentypen

Diskrete Variablen

Diskrete (kategoriale) Variablen sind Variablen, die nur eine endliche Anzahl an verschiedenen Werten annehmen und meist durch ganze Zahlen repräsentiert sind.

Kontinuierliche Variablen

Kontinuierliche Variablen sind Variablen, die unendlich viele Werte annehmen können und meist durch die reellen Zahlen repräsentiert sind.

Einordnung einer Variable als diskret oder kontinuierlich ist eine Modellierungsannahme

Geschlecht	m/w vs. m/w/d vs. Kontinuum
Alter	Zeit als reelle Zahl vs. 20, 21, 22, ..., 100
Reaktionszeiten	Zeit als reelle Zahl vs. floating point numbers

Allgemeine Systematik von Studiendesigns

Randomisierte kontrollierte Studie (Experiment)

- Die experimentellen Einheiten werden den Versuchsbedingungen zufällig zugeordnet
- Beispiel: Online Psychotherapie vs. Klassische Psychotherapie bei Depression

Nicht-randomisierte kontrollierte Studie (Quasiexperiment)

- Untersuchung natürlich bzw. bereits bestehender Gruppen
- Beispiel: Online Psychotherapie bei Depression vs. Schizophrenie

Analyse eines bestehenden Datensatzes (Korrelationsstudie)

- Nicht-randomisierte, nicht kontrollierte Studie
- Beobachtungsstudie ohne Intervention
- Beispiel: Analyse von Paneldaten

Charakteristika randomisierter kontrollierter Studien

- Vorhandensein einer kausaltheoretischen Hypothese vor Versuchsbeginn
- Gute Manipulierbarkeit von unabhängigen Variablen
- Explizite Operationalisierung der untersuchten Konstrukte
- Kontrollierbarkeit möglichst vieler Versuchsbedingungen
- Typisch für bereits gut erschlossene Gegenstandsbereiche

Faktorielle Studiendesigns

- Kategoriale unabhängige Variable, die Faktor genannt wird
- Die Werte der unabhängigen Variablen werden Level genannt
- Einfaktorielle oder mehrfaktoriell

Parametrische Studiendesigns

- Kontinuierliche unabhängige Variable
- Die Werte der unabhängigen Variablen werden oft Level genannt
- Meist einfaktoriell

Between-Group Designs | Studiendesigns mit Randomisierung

- Gesamtgruppe wird zufällig auf experimentelle Bedingungen aufgeteilt
- Einfaktoriell, mehrfaktoriell
- Häufig in der klinischen Forschung verwendet

Within-Group Designs | Studiendesigns mit Wiederholungsmessung

- Eine Gesamtgruppe wird sämtlichen experimentellen Bedingungen unterzogen
- Einfaktoriell, mehrfaktoriell, parametrisch
- Häufig in der psychologischen Grundlagenforschung verwendet

Block Designs | Studiendesigns mit parallelisierten Gruppen

- Gesamtgruppe wird gesteuert auf experimentelle Bedingungen aufgeteilt
- Experimentelle Gruppen werden anhand bestimmter Merkmale parallelisiert
- Häufiger in der klinischen Forschung verwendet

Mixed Designs | Mischdesigns

- Mischungen aller obiger Studiendesignstypen

Randomisierung bei Between-Group Designs

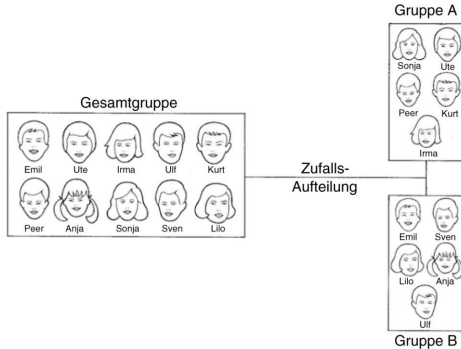


Abbildung 5.1: Zufallsaufteilung einer Gesamtgruppe von N= 10 Probanden (links) in zwei gleiche große Untergruppen (rechts). (Aus Sarris, 1999)

Designschemata

- R: Randomisierung
- O: Observation (Test, Messung)
- X: Exposition experimenteller Bedingung
- Experimentelle Bedingungen von oben nach unten
- Zeitliche Abfolge von links nach rechts

Beispiel

R	X	O
R		O

- Bedingungszuweisung erfolgt durch Randomisierung
- Nur eine Gruppe erhält das Treatment
- Beide Gruppen absolvieren die Messung

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Randomisierte einfaktorielle Studiendesigns

- Gesamtgruppe wird zufällig auf experimentelle Bedingungen aufgeteilt
- Eine unabhängige Variable mit zwei oder mehr Leveln
- Populäres Designs in der klinischen Forschung
- Varianten
 - Treatment- und Kontrollgruppe
 - Treatment- und Placebogruppe
 - Zwei Treatmentgruppen (und Kontrollgruppe)
 - Pretest-Posttest Designs

No-Treatment Kontrollgruppe

R	X	O
R		O

- Vergleich eines Treatments zu keinem Treatment

Placebo Kontrollgruppe

R	X	O
R	X_P	O

- Placebo = Scheintreatment
- Vergleich eines Treatments zu keinem Treatment
- Kontrolle studieninduzierter Effekte (Placeboeffekte)

Vergleich zweier Treatments

R	X_A	O
R	X_B	O

- Vergleich Standardtreatment A und neues Treatment B
- Keine Aussage über Effektivität des Standardtreatments

Zwei-Treatment Vergleich mit Placebo-Kontrollgruppe

R	X_A	O
R	X_B	O
R	X_P	O

- Vergleich Standardtreatment A und neues Treatment B
- Aussage über Effektivität des Standardtreatments möglich
- Placebotreatment kann ethisch nicht vertretbar sein

Beispiel: Einfluss von Psychotherapie auf Depressionssymptomatik

- Klassische Psychotherapie (A)
- Online Psychotherapie (B)
- Seelsorge (P)

→ Keine Aussagen über Pre-Treatment Gruppenunterschiede möglich

→ Keine Aussage über Dropout Charakteristika möglich

Pre-Posttest Designs

R	O	X_A	O
R	O	X_B	O
R	O		O

- Fokus auf Treatment-induzierte Verbesserungen/Verschlechterungen
- Subtraktion von Pre-Test-Gruppenunterschieden möglich
- Untersuchung von Dropout Charakteristika möglich
- Mögliches Auftreten von Testeffekten (Lernen, Gewöhnung, Ermüdung)
- Höherer Zeit- und Kostenaufwand

Randomisierte einfaktorielle Studiendesigns

Beispiel: Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung
(Gruppen von $n = 12$)

Psychotherapie

Klassisch

Prä-BDI



Post-BDI

Online

Prä-BDI



Post-BDI

⇒ Randomisiertes einfaktorielles Preposttest Design ohne Kontrollgruppe

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Mehrfaktorielle Studiendesigns

- Kombination mehrerer experimenteller Faktoren in einem Studiendesign

Crossed Design

- Jedes Level jedes Faktors wird mit allen Leveln aller Faktoren kombiniert.

Nested Design

- Einige Level eines Faktors werden nicht mit allen anderen Faktorleveln kombiniert.

⇒ Prototypisch sind zweifaktorielle Studiendesigns mit crossed design

Randomisiertes zweifaktorieller Studiendesigns mit crossed design

- Eine univariate abhängige Variable bestimmt an individuellen experimentellen Einheiten.
- Zwei diskrete unabhängige Variablen, die mindestens zweistufig sind.
- Die unabhängigen Variablen werden *Faktoren* genannt.
- Die Stufen der Faktoren werden *Level* genannt.
- Jedes Level eines Faktors wird mit allen Level des anderen Faktors kombiniert

Zweifaktorielle Studiendesigns werden üblicherweise anhand ihrer Faktorlevel bezeichnet

2 × 2 Design: Faktor A mit Level 1,2	Faktor B mit Level 1,2
2 × 3 Design: Faktor A mit Level 1,2	Faktor B mit Level 1,2,3
4 × 2 Design: Faktor A mit Level 1,2,3,4	Faktor B mit Level 1,2
3 × 1 Design: Faktor A mit Level 1,2,3	Faktor B mit Level 1

- 2 × 2 Studiendesigns sind sehr populär, wir fokussieren auf diesen Fall.

⇒ Das entsprechende datenanalytische Verfahren ist die Varianzanalyse (ANOVA)

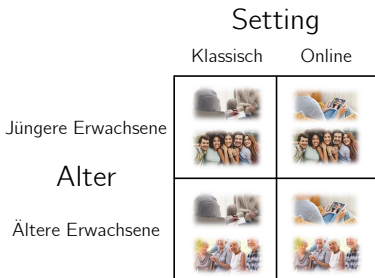
Konzeptuelles Design eines 2 x 2 Studiendesigns

		Faktor B	
		Level 1	Level 2
Faktor A	Level 1	A1B1	A1B2
	Level 2	A2B1	A2B2

Randomisierte mehrfaktorielle Studiendesigns

Anwendungsbeispiel eines randomisierten 2 x 2 Studiendesigns

- Ist Psychotherapie bei Depression im klassischen oder im Online Setting wirksamer?
- Ist Psychotherapie bei Depression bei jüngeren oder älteren Patient:innen wirksamer?



- Ist die Wirksamkeit der Psychotherapie vom Setting abhängig?
- Ist die Wirksamkeit der Psychotherapie vom Alter der Patient:innen abhängig?

Verlaufsschema eines randomisierten 2×2 Studiendesigns

R	X_{A1B1}	O
R	X_{A1B2}	O
R	X_{A2B1}	O
R	X_{A2B2}	O

- Pre-Posttest Designs möglich
- Placebo Kontrollgruppen möglich

Randomisierte mehrfaktorielle Studiendesigns

Daten bei 2 x 2 Studiendesigns

		Faktor B	
		Level 1	Level 2
Faktor A	Level 1	1.2	4.1
		0.4	1.8
		⋮	⋮
		2.1	5.9
	Level 2	0.1	7.4
		2.7	9.5
		⋮	⋮
		1.4	6.1

Haupteffekte und Interaktionen

Hinsichtlich der Gruppenmittelwerte bei 2×2 Studiendesigns unterscheidet man **Haupteffekte** und **Interaktionen**

- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor A*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor A, jeweils gemittelt über die zwei Level von Faktor B, unterscheiden.
- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor B*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor B, jeweils gemittelt über die zwei Level von Faktor A, unterscheiden.
- Intuitiv spricht man vom Vorliegen einer *Interaktion der Faktoren A und B*, wenn der Unterschied der Gruppenmittelwerte von Faktor A zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor B ausgeprägt ist bzw. wenn der Unterschied der Gruppenmittelwerte von Faktor B zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor A ausgeprägt ist.

Intuitiv beziehen sich Haupteffekte also auf (marginale) Unterschiede (Differenzen), während sich Interaktionen auf Unterschiede von Unterschieden (Differenzen von Differenzen) beziehen.

Das Vorhandensein einer Interaktion besagt lediglich, dass sich die Unterschiede der Gruppenmittelwerte zwischen den Leveln eines experimentellen Faktors in Abhängigkeit von den Leveln des anderen experimentellen Faktors ändern, es macht aber keine Aussage darüber, warum dies so ist. Haupteffekte und Interaktionen sind lediglich Datenmuster, keine mechanistischen wissenschaftlichen Theorien.

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Welche Therapieform ist bei Depression wirksamer?

Online Psychotherapie



Klassische Psychotherapie



→ Klinische Psychologie, Klinische Diagnostik, MSc Psychotherapie

Wahrscheinlichkeitstheorie und Frequentistische Inferenz WiSe 21/22 (1) Einführung

Anwendungskontext

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Becks Depressions-Inventar (BDI) zur Depressionsdiagnostik

BCI-II		Fragebogen	
Name		Alter	Jahreszahl
			mit / ab
Achtung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede Einzel-Gruppe von Aussagen sorgfältig durch und suchen Sie sich -Eine in jeder Gruppe eine Aussage heraus. Sie am besten beachtet, wie Sie sich die Meiste oder Mehrere, eindeutigste, größte, höchste Konkretheit die für Sie selber Aussage an, die Sie sich am zutreffendsten haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichwertig auf Sie zutreffen, können Sie die Aussage mit der höchsten Zahl aus. Jedes No teile darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, die gilt auch für Gruppe 19 (Vorstellungen & Schlagschlag) oder Gruppe 18 (Veränderungen des Appetits).			
1.) Traurigkeit	<input type="checkbox"/> Ich bin nicht traurig. <input type="checkbox"/> Ich bin oft traurig. <input type="checkbox"/> Ich bin ständig traurig. <input type="checkbox"/> Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.	4.) Bestrafungsgefühle	<input type="checkbox"/> Ich habe nicht die Gefühl, für etwas bestraft zu sein. <input type="checkbox"/> Ich habe das Gefühl, vielleicht bestraft zu werden. <input type="checkbox"/> Ich empfinde, bestraft zu werden. <input type="checkbox"/> Ich habe das Gefühl, bestraft zu sein.
2.) pessimismus	<input type="checkbox"/> Ich sehe nicht mehr in die Zukunft. <input type="checkbox"/> Ich sehe trübsamer in die Zukunft als sonst. <input type="checkbox"/> Ich bin trübsam und erwarte nicht, dass meine Situation besser wird. <input type="checkbox"/> Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.	7.) Selbstabwertung	<input type="checkbox"/> Ich fühle mich nur geduldet und wie etwas. <input type="checkbox"/> Ich habe Vorurteile in mich wertlos. <input type="checkbox"/> Ich bin von mir enttäuscht. <input type="checkbox"/> Ich lehne mich völlig ab.
3.) Versagensgefühle	<input type="checkbox"/> Ich fühle mich nicht als Versager. <input type="checkbox"/> Ich habe häufiger Versagensgefühle. <input type="checkbox"/> Wenn ich zurechtbilde, sehe ich eine Menge Fortschritte. <input type="checkbox"/> Ich habe das Gefühl, als Mensch ein völlig Versager zu sein.	8.) Selbstvorwürfe	<input type="checkbox"/> Ich kritisiere oder tadelte mich nicht mit dem ich sonst. <input type="checkbox"/> Ich bin mir gegenüber kritischer als sonst. <input type="checkbox"/> Ich kritisiere mich für all meine kleinen Mängel. <input type="checkbox"/> Ich gebe mir die Schuld für alles Schlechte, was passiert.
4.) Verlust von Freude	<input type="checkbox"/> Ich kann die Dinge genauso gut genießen wie früher. <input type="checkbox"/> Ich kann die Dinge nicht mehr so genießen wie früher. <input type="checkbox"/> Dinge, die mir früher Freude gemacht haben, kann ich kaum mehr genießen. <input type="checkbox"/> Dinge, die mir früher Freude gemacht haben, kann ich überhaupt nicht mehr genießen.	9.) Selbstmordgedanken	<input type="checkbox"/> Ich denke nicht daran, mir etwas anzutun. <input type="checkbox"/> Ich denke manchmal an Selbstmord. <input type="checkbox"/> Ich möchte mich an liebsten umbringen. <input type="checkbox"/> Ich möchte mich anbringen, wenn ich die Selbstdenken dazu hätte.
5.) Schuldgefühle	<input type="checkbox"/> Ich habe keine besonderen Schuldgefühle. <input type="checkbox"/> Ich habe viele besonderen Schuldgefühle, die ich gerne habe oder hätte tun sollen. <input type="checkbox"/> Ich habe die meisten Züge Schuldgefühle. <input type="checkbox"/> Ich habe ständig Schuldgefühle.	10.) Weisheit	<input type="checkbox"/> Ich weine nicht öfter als früher. <input type="checkbox"/> Ich weine jetzt mehr als früher. <input type="checkbox"/> Ich weine beim geringsten Anlass. <input type="checkbox"/> Ich weine gern weinen, aber ich kann nicht.

11.) Unsicherheit	<input type="checkbox"/> Ich bin nicht unruhiger als sonst. <input type="checkbox"/> Ich bin unruhiger als sonst. <input type="checkbox"/> Ich bin so unruhig, dass es mir schwerfällt, still zu sitzen. <input type="checkbox"/> Ich bin so unruhig, dass ich mich ständig bewegen oder etwas tun muss.	12.) Interessenverlust	<input type="checkbox"/> Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren. <input type="checkbox"/> Ich habe weniger Interesse an anderen Menschen oder an Dingen als sonst. <input type="checkbox"/> Ich habe das Interesse an anderen Menschen oder Dingen zum größten Teil verloren. <input type="checkbox"/> Zu fällt mir schwer, mich überhaupt für irgend etwas zu interessieren.
13.) Entscheidungsfähigkeit	<input type="checkbox"/> Ich bin so entscheidungsfähig wie immer. <input type="checkbox"/> Ich fühle mich schwerer als sonst, Entscheidungen zu treffen. <input type="checkbox"/> Es fällt mir sehr viel schwerer als sonst, Entscheidungen zu treffen. <input type="checkbox"/> Ich habe Mühe, überhaupt Entscheidungen zu treffen.	14.) Wertlosigkeit	<input type="checkbox"/> Ich fühle mich nicht wertlos. <input type="checkbox"/> Ich habe mich für weniger wertvoll und nützlich als sonst. <input type="checkbox"/> Vergleichen mit anderen Menschen fühle ich mich viel weniger wert. <input type="checkbox"/> Ich fühle mich völlig wertlos.
15.) Energieverlust	<input type="checkbox"/> Ich habe so viel Energie wie immer. <input type="checkbox"/> Ich habe weniger Energie als sonst. <input type="checkbox"/> Ich habe so wenig Energie, dass ich kaum noch etwas schaffe. <input type="checkbox"/> Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.	16.) Veränderungen der Schlafgewohnheiten	<input type="checkbox"/> Meine Schlafgewohnheiten haben sich nicht verändert. <input type="checkbox"/> Ich schlafe etwas mehr als sonst. <input type="checkbox"/> Ich schlafe viel mehr als sonst. <input type="checkbox"/> Ich schlafe viel weniger als sonst. <input type="checkbox"/> Ich schlafe fast den ganzen Tag. <input type="checkbox"/> Ich weiche 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.
17.) Reizbarkeit	<input type="checkbox"/> Ich bin nicht reizbarer als sonst. <input type="checkbox"/> Mein Appetit hat sich nicht verändert. <input type="checkbox"/> Mein Appetit ist etwas schlechter als sonst. <input type="checkbox"/> Mein Appetit ist etwas größer als sonst. <input type="checkbox"/> Mein Appetit ist viel schlechter als sonst. <input type="checkbox"/> Mein Appetit ist viel größer als sonst. <input type="checkbox"/> Ich habe überhaupt keinen Appetit. <input type="checkbox"/> Ich habe ständig Heißhunger.	18.) Veränderungen des Appetits	<input type="checkbox"/> Ich kann mich so gut konzentrieren wie immer. <input type="checkbox"/> Ich kann mich nicht mehr so gut konzentrieren wie sonst. <input type="checkbox"/> Zu fällt mir schwer, mich längere Zeit auf irgend etwas zu konzentrieren. <input type="checkbox"/> Ich kann mich überhaupt nicht mehr konzentrieren.
19.) Ermüdung oder Erschöpfung	<input type="checkbox"/> Ich fühle mich nicht müde oder erschöpfter als sonst. <input type="checkbox"/> Ich werde schneller müde oder erschöpft als sonst. <input type="checkbox"/> Für viele Dinge, die ich üblicherweise tue, bin ich zu müde oder erschöpft. <input type="checkbox"/> Ich kann mich überhaupt nicht mehr konzentrieren.	20.) Verlust an sexuellem Interesse	<input type="checkbox"/> Mein Interesse an Sexualität hat sich in letzter Zeit nicht verändert. <input type="checkbox"/> Ich interessiere mich weniger für Sexualität als früher. <input type="checkbox"/> Ich interessiere mich jetzt viel weniger für Sexualität. <input type="checkbox"/> Ich habe das Interesse an Sexualität völlig verloren.

0 - 8 keine Depression

9 - 13 minimale Depression

14 - 19 leichte Depression

20 - 28 mittelschwere Depression

29 - 63 schwere Depression

Wahrscheinlichkeitstheorie und Frequentistische Inferenz WiSe 21/22 (1) Einführung

Evidenzbasierte Evaluation von Psychotherapieformen bei Depression - Verhaltensdaten

Patient ID	Bedingung	Prae-PT-BDI	Post-PT-BDI
1	Online	29	19
2	Online	27	30
3	Klassisch	16	11
4	Online	19	26
5	Online	35	29
6	Klassisch	48	10
7	Online	20	15
8	Online	18	11
9	Online	48	28
10	Klassisch	49	23
11	Klassisch	39	10
12	Online	36	24
13	Online	21	27
14	Online	46	14
15	Online	20	17
16	Klassisch	23	10

Wahrscheinlichkeitstheorie und Frequentistische Inferenz WiSe 21/22 (1) Einführung



Digital Interventions for the Treatment of Depression: A Meta-Analytic Review

Isaac Moshe¹, Yannik Terhorst^{2, 3}, Paula Philippi³, Matthias Domhardt³, Pim Cuijpers⁴, Ioana Cristea⁵,
Laura Pulkki-Räback^{1, 6}, Harald Baumeister³, and Lasse B. Sander⁷

Abstract

The high global prevalence of depression, together with the recent acceleration of remote care owing to the COVID-19 pandemic, has prompted increased interest in the efficacy of digital interventions for the treatment of depression. We provide a summary of the latest evidence base for digital interventions in the treatment of depression based on the largest study sample to date. A systematic literature search identified 83 studies ($N = 15,530$) that randomly allocated participants to a digital intervention for depression versus an active or inactive control condition. Overall heterogeneity was very high ($I^2 = 84\%$). Using a random-effects multilevel metaregression model, we found a significant medium overall effect size of digital interventions compared with all control conditions ($g = .52$). Subgroup analyses revealed significant differences between interventions and different control conditions (WLC: $g = .70$; attention: $g = .36$; TAU: $g = .31$), significantly higher effect sizes in interventions that involved human therapeutic guidance ($g = .63$) compared with self-help interventions ($g = .34$), and significantly lower effect sizes for effectiveness trials ($g = .30$) compared with efficacy trials ($g = .59$). We found no significant difference in outcomes between smartphone-based apps and computer- and Internet-based interventions and no significant difference between human-guided digital interventions and face-to-face psychotherapy for depression, although the number of studies in both comparisons was low. Findings from the current meta-analysis provide evidence for the efficacy and effectiveness of digital interventions for the treatment of depression for a variety of populations. However, reported effect sizes may be exaggerated because of publication bias, and compliance with digital interventions outside of highly controlled settings remains a significant challenge.

Moshe et al. (2021)



Digital Interventions for the Treatment of Depression: A Meta-Analytic Review

Isaac Moshe¹, Yannik Terhorst^{2, 3}, Paula Philippi³, Matthias Domhardt³, Pim Cuijpers⁴, Ioana Cristea⁵,
Laura Pulkki-Råback^{1, 6}, Harald Baumeister³, and Lasse B. Sander⁷

Results

Is There a Difference in Effect Size Between Digital Interventions and Face-to-Face Therapy?

Only three comparisons from three studies were available comparing digital interventions and individual face-to-face therapy, all of which involved interventions with human guidance. The comparison adjusted for baseline differences indicated there was a nonsignificant difference to face-to-face therapy of $g = -.01$, 95% CI [-2.73, 2.72], $p = .982$; $I^2 < .001$, 95% CI [.00, 100]; $\sigma^2_{within} = \text{fixed}$, $\sigma^2_{between} < .001$, 95% CI [.000, 9.195]. For group face-to-face therapy we identified only two studies, which provided a total of three data points at post. Again, based on this limited evidence, there was no significant difference (unadjusted for baseline difference, owing to convergence; $g = .17$, 95% CI [-2.91, 3.26], $p = .609$); $I^2 = 69$, 95% CI [.00, 100]; $\sigma^2_{within} \leq .001$, 95% CI [.000, .976], $\sigma^2_{between} = .093$, 95% CI [.000, > 10.000].

Moshe et al. (2021)



Digital Interventions for the Treatment of Depression: A Meta-Analytic Review

Isaac Moshe¹, Yannik Terhorst^{2, 3}, Paula Philippi³, Matthias Domhardt³, Pim Cuijpers⁴, Ioana Cristea⁵,
Laura Pulkki-Råback^{1, 6}, Harald Baumeister³, and Lasse B. Sander⁷

Discussion

The current review identified only three studies that directly compared digital interventions with face-to-face therapy. We found no significant difference in outcomes between the two conditions, supporting findings from [Carlbring et al. \(2018\)](#). However, unlike the study by Carlbring and colleagues, which mixed individual and group-based psychotherapy, our analysis was limited to individual therapy alone. It is important to note that these were highly controlled trials across multiple conditions and with low sample sizes, the majority of whom were self-referred. Moreover, participants needed to consent to both possible assignments of Internet or on-site care, likely leading to a highly selective subsample of depressed participants in need of mental health care. Based on these findings, and the notable lack of high-quality studies providing direct comparisons, we believe it is premature to conclude that digital interventions are as equally effective as face-to-face psychotherapy for the treatment of depression and mark this out as a critical area for future research.

Moshe et al. (2021)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn^h, Andreas Maercker^h

Abstract

Background and aims: In the past decade, a large body of research has demonstrated that internet-based interventions can have beneficial effects on depression. However, only a few clinical trials have compared internet-based depression therapy with an equivalent face-to-face treatment. The primary aim of this study was to compare treatment outcomes of an internet-based intervention with a face-to-face intervention for depression in a randomized non-inferiority trial.

Method: A total of 62 participants suffering from depression were randomly assigned to the therapist-supported internet-based intervention group ($n=32$) and to the face-to-face intervention ($n=30$). The 8 week interventions were based on cognitive-behavioral therapy principles. Patients in both groups received the same treatment modules in the same chronological order and time-frame. Primary outcome measure was the Beck Depression Inventory-II (BDI-II); secondary outcome variables were suicidal ideation, anxiety, hopelessness and automatic thoughts.

Results: The intention-to-treat analysis yielded no significant between-group difference (online vs. face-to-face group) for any of the pre- to post-treatment measurements. At post-treatment both treatment conditions revealed significant symptom changes compared to before the intervention. Within group effect sizes for depression in the online group ($d=1.27$) and the face-to-face group ($d=1.37$) can be considered large. At 3-month follow-up, results in the online group remained stable. In contrast to this, participants in the face-to-face group showed significantly worsened depressive symptoms three months after termination of treatment ($t=-2.05$, $df=19$, $p<.05$).

Limitations: Due to the small sample size, it will be important to evaluate these outcomes in adequately-powered trials.

Conclusions: This study shows that an internet-based intervention for depression is equally beneficial to regular face-to-face therapy. However, more long term efficacy, indicated by continued symptom reduction three months after treatment, could be only be found for the online group.

Wagner, Horn, and Maercker (2014)



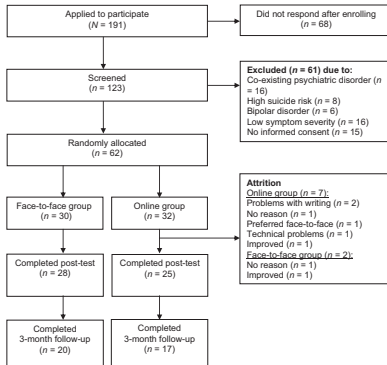
Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn^h, Andreas Maercker^h

Participants



Wagner, Horn, and Maercker (2014)



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn^h, Andreas Maercker^h

Demographics

Table 1

Demographic and descriptive characteristics of the online and face-to-face groups at baseline.

	Online group (n=32)	Face-to-face group (n=30)
Age (M, SD)	37.25 (11.41)	38.73 (11.41)
Age range	20–67	19–62
Gender (% female)	78%	50%
Educational level (%)		
Primary education	16	27
Secondary education	37	27
University	47	47
Marital status (%)		
Single	59	57
Partnership/married	19	20
Divorced	16	7
Widowed	6	17
Professional status (%)		
Full-time work	72	67
Sick leave	3	10
Unemployed	22	20
Retired	0	3
No current antidepressants (%)	91	70
BDI-score, pre-test (M, SD)	22.96 (6.07)	23.41 (7.63)
BDI pre-test (11–17) in %	16.7	23.3
BDI pre-test (18–29) in %	63.3	56.7
BDI pre-test (≥30) in %	20	20
Completer in %	78	93

Wagner, Horn, and Maercker (2014)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn^h, Andreas Maercker^h

Procedure

2.3. Procedure

Primary and secondary outcome measures were collected at pre-treatment, post-treatment and 3-month follow-up. All measures for both intervention groups were administered through online diagnostics. A number of studies have shown that online format questionnaires produce results as valid as pen-and-paper questionnaires (Fidy, 2008; Hollandare et al., 2010). The 62 applicants included in the study were randomized by a true random-number service (<http://www.random.org>) using a 1:1 ratio, with 32 participants randomly allocated to the online group and 30 to the face-to-face treatment group. Randomization was performed by the study coordinator and was not stratified by any participant characteristics.

Wagner, Horn, and Maercker (2014)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn^h, Andreas Maercker^b

Interventions

2.4. Interventions

Online and face-to-face intervention groups received a brief (8 weeks) cognitive-behavioral therapy (CBT) program for depression (Hautzinger, 2003). This German manual is based on the cognitive theory of depression of Beck and colleagues (Hautzinger et al., 2006). The program involved the following CBT modules: (1) introduction, (2) behavioral analysis, (2) planning of activities, (3) daily structure, (4) life review, (5) cognitive restructuring, (6) social competence, and (7) relapse prevention. The life-review module at the mid-treatment time-point aimed to encourage participants to revisit past experiences and to activate positive memories and individual resources in order to achieve a balance between positive and negative memories (Preschl et al., 2012). Further, patients in both groups were given the same psychoeducation and received the treatment modules in the same chronological order. Patients in the face-to-face condition attended one-hour weekly treatment sessions for 8 weeks with their allocated psychologist in the Department of Psychopathology and Clinical Intervention at the University of Zurich. They were also given weekly homework assignments (e.g., daily structure diaries, negative thoughts log).

The online intervention was given as a guided intervention with intensive therapist contact, based on the principles applied in a number of previous studies (Lange et al., 2003; Ruwaard et al., 2007; Ruwaard et al., 2009; Wagner et al., 2006). The internet-based treatment manual was derived from the same cognitive-behavioral treatment modules for depression as the face-to-face intervention (Hautzinger, 2003). The therapist time involved responding to texts, requiring 20–50 min per text, depending on the therapist's experience with internet-based therapies.

Each scheduled writing assignment lasted 45 min and patients were given two writing assignments in each week of the 8-week treatment period. Therapists provided individual written feedback within one working day, along with instructions for the next writing assignment. Model responses for the therapists were available, but they also had the option to provide their own commentary or supportive feedback on their patients' texts.

Wagner, Horn, and Maercker (2014)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn[†], Andreas Maercker[‡]

Outcome measure (Abhängige Variable)

2.6. Outcome measures

All outcome measures were collected at pre-treatment, post-treatment and 3-month follow-up.

2.7. Primary outcome measure

The primary outcome measure of this study was depression assessed with the German version (Hautzinger et al., 2006) of the Beck Depression Inventory-II (BDI; Beck et al., 1996), comprised of 21 multiple-choice items assessing specific symptoms of depression. Symptom severity was defined for mild or moderate depression (BDI score: 11–17); moderate to severe depression (BDI score: 18–29); and severe depression (BDI score ≥ 30) (Hautzinger et al., 2006). Recovery was defined as BDI-II at post-treatment measurement of ≤ 10 .

Wagner, Horn, and Maercker (2014)



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn[†], Andreas Maercker[‡]

Results

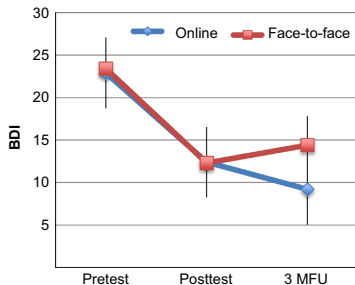


Fig. 2. Online intervention in comparison to a face-to-face group measured with the Beck Depression Inventory (BDI-II) at pretest, posttest and 3-months-follow-up, including standard error.

Wagner, Horn, and Maercker (2014)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn[†], Andreas Maercker[‡]

Discussion

However, analysis revealed that from post-treatment to 3-month follow-up a difference between the internet-based intervention and the face-to-face group could be found. Symptom reductions were maintained for all primary and secondary outcomes for the online group three months after treatment. In contrast to this, participants in the face-to-face group significantly worsened from post-treatment to 3-month follow-up in terms of depressive symptoms. Further, significant differences were found for the face-to-face group from post-treatment to the 3-month follow-up for symptoms of anxiety and automatic negative thoughts and a nearly significant effect was observed for depression. Altogether it appears that the treatment effects from pre-treatment to 3-month follow-up were larger in the online group than in the face-to-face group. Moreover, at the 3-month follow-up more participants in the online group indicated clinically significant changes than in the face-to-face group. Reasons for this might include that the online intervention has less personal guidance and therefore puts a stronger focus on self-responsibility to conduct the treatment modules and homework assignments than the face-to-face intervention. This might evoke a stronger, longer-lasting sense of self-efficacy in handling negative thoughts and depressive behavior. Further, no significant difference could be

Wagner, Horn, and Maercker (2014)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{a,*}, Andrea B. Horn^b, Andreas Maercker^b

Conclusion

Depression has become a very prevalent und costly disorder and in most countries therapeutic services do not manage to meet the needs presented by this growing demand. This trial gives preliminary results that a brief internet-based intervention for depression is as effective as comparable face-to-face interventions. Internet-based intervention may be the solution for tackling this epidemic in a more cost-effective way than traditional face-to-face therapies. However, further research is needed to replicate these findings and possible differences in underlying mechanisms between online and face-to-face interventions need to be evaluated.

Wagner, Horn, and Maercker (2014)



Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn[†], Andreas Maercker[‡]



Australian New Zealand Clinical Trials Registry
The ANZCTR is an online registry of clinical trials being undertaken in Australia, New Zealand and elsewhere.

[FAQS](#) | [HOW TO GET INVOLVED](#) | [NEWS](#) | [ABOUT US](#)

 [CREATE ACCOUNT](#)  [LOGIN](#)

Registering a new trial? To achieve prospective registration, we recommend submitting your trial for registration at the same time as ethics submission.

[SEARCH COVID-19 STUDIES](#)

[COVID-19 RESOURCES](#)

Searching for a trial?

[SEARCH NOW](#)

The search function allows you to conduct either a basic search or an advanced search of clinical trials available on the ANZCTR database. Once you find a relevant trial, you will be able to contact the person listed as the 'public contact' on the trial record for more information.

Want to register a trial?

[REGISTER YOUR TRIAL](#)

In order to fulfil the prospective registration requirement, i.e. obtain registration number (ACTRN) prior to enrolment of the first participant, we recommend that Australian and New Zealand registrants commence the registration process at least eight weeks prior to the anticipated recruitment start date. All other registrants should allow substantially longer.

Contents lists available at ScienceDirect



Journal of Affective Disorders

Journal homepage: www.elsevier.com/locate/jad



Research report

Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial[☆]



Birgit Wagner^{☆*}, Andrea B. Horn^h, Andreas Maercker^h



Australian New Zealand Clinical Trials Registry

The ANZCTR is an online registry of clinical trials being undertaken in Australia, New Zealand and elsewhere.

FAQ | HOW TO GET INVOLVED | NEWS | ABOUT US

CREATE ACCOUNT | LOGIN

Trial Search

ACTRN12611000563865



The safety and scientific validity of each study registered on the ANZCTR is the responsibility of the study sponsor and investigator. Listing a study on the ANZCTR does not mean it has been endorsed by the ANZCTR. Before participating in a study, talk to your health care provider and refer to [information for consumers](#). For additional information regarding COVID-19 studies, please click [here](#).

Advanced search options

Map of recruitment locations

Number of records found: 1

Download ALL ANZCTR hits to Excel

DOWNLOAD

HTML downloads of search results are currently unavailable. Please use the Excel download function instead.

Page: 1

Trial from ANZCTR

[View full record](#) Select for download

In male and female adults suffering from depression, is online cognitive behavioural therapy more effective than face-to-face treatment in terms of clinical outcomes and working alliance?

[Active, not recruiting](#) [Ethics status: Approved](#) [Retrospectively registered](#) [Not up to date \(last updated: 14/6/2018\)](#)

- Key trial information
- Eligibility
- Contact details and further information

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Anwendungsbeispiel

Experimentelle Bedingung
(Gruppen von $n = 40$)

Psychotherapie

Face-to-face

Prä-BDI



Post-BDI

Online

Prä-BDI



Post-BDI

Waitlist Control

Prä-BDI

Post-BDI

Anwendungsbeispiel

Experimentelle Einheiten

Randomisierte Zuordnung von Patient:innen zu den experimentellen Gruppen

- Unabhängig und identisch verteilte Fehlervariablen

Abhängige Variable

Negative Post-BDI vs. Prä-BDI Differenz

- Univariate abhängige Variable \Leftrightarrow Ein Wert pro Patient:in
- Positive Werte \Leftrightarrow Verringerung der Depressionssymptomatik
- Negative Werte \Leftrightarrow Verstärkung der Depressionssymptomatik

Unabhängige Variablen

Experimentelle Gruppen \Leftrightarrow Psychotherapie Setting

- Face-to-face (F2F), Online (ONL), Waitlist Control (WLC)

Patient:innen Alter in Jahren

Anzahl Therapiestunden

Beispieldatensatz

	A	B	C	D	E	F	G
1	ID	Condition	PreBDI	PostBDI	BDI	Age	Duration
2	1	F2F	17	62	-45	38	13
3	2	F2F	23	31	-8	55	16
4	3	F2F	36	31	5	75	22
5	4	F2F	57	11	46	29	15
6	5	F2F	13	48	-35	45	19
7	6	F2F	57	29	28	33	22
8	7	F2F	60	32	28	46	13
9	8	F2F	42	13	29	28	20
10	9	F2F	40	14	26	48	20
11	10	F2F	4	38	-34	77	18
12
13	41	ONL	52	18	34	22	20
14	42	ONL	41	56	-15	52	14
15	43	ONL	49	28	21	73	23
16	44	ONL	35	49	-14	42	23
17	45	ONL	33	55	-22	23	23
18	46	ONL	50	26	24	28	21
19	47	ONL	1	4	-3	39	16
20	48	ONL	30	21	9	29	21
21	49	ONL	46	46	0	28	12
22	50	ONL	44	21	23	33	23
23
24	81	WLC	27	17	10	60	12
25	82	WLC	45	14	31	39	16
26	83	WLC	25	33	-8	44	18
27	84	WLC	20	17	3	80	19
28	85	WLC	48	11	37	71	15
29	86	WLC	13	33	-20	77	17
30	87	WLC	45	35	10	69	16
31	88	WLC	8	8	0	67	23
32	89	WLC	15	16	-1	36	13
33	90	WLC	9	45	-36	66	13
34
35	120	WLC	40	29	11	48	20

Grundbegriffe

Randomisierte einfaktorielle Studiendesigns

Randomisierte mehrfaktorielle Studiendesigns

Anwendungskontext

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die Begriffe der empirischen Studie, der theoretischen Studie, und der Methodenstudie.
2. Erläutern Sie die Begriffe der unabhängigen Variable, der abhängigen Variable, und der experimentellen Einheit.
3. Erläutern Sie die Begriffe der diskreten Variable und der kontinuierlichen Variable.
4. Erläutern Sie die Begriffe der randomisierten und der nicht-randomisierten kontrollierten Studie.
5. Erläutern Sie die Begriffe des Quasiexperiments und der Korrelationsstudie.
6. Nennen Sie drei Charakteristika randomisierter kontrollierter Studien.
7. Erläutern Sie die Begriffe des faktoriellen und des parametrischen Studiendesigns.
8. Erläutern Sie die Begriffe des Between-Group Designs und des Within-Group Designs.
9. Erläutern Sie die Begriffe des Studiendesigns mit Randomisierung bzw. mit Wiederholungsmessung.
10. Erläutern Sie den Begriff des randomisierten einfaktoriellen Studiendesigns.
11. Diskutieren Sie Vor- und Nachteile von No-Treatment und Placebo-Treatment Kontrollgruppen.
12. Diskutieren Sie Vor- und Nachteile von Zwei-Treatment Vergleichen ohne und mit Placebo-Kontrollgruppe.
13. Erläutern Sie Vor- und Nachteile von reinen Posttest-Designs und Pre- und Posttest Designs.
14. Erläutern Sie die Begriffe des mehrfaktoriellen Studiendesigns, des Crossed Designs, und des Nested Designs.
15. Erläutern Sie den Begriff des randomisierten zweifaktoriellen Studiendesigns mit Crossed Design.
16. Wieviele Faktoren mit jeweils wie vielen Leveln hat ein $3 \times 4 \times 2$ Design?
17. Wieviele experimentelle Bedingungen hat ein $3 \times 4 \times 2$ Design?
18. Erläutern Sie die Begriffe des Haupteffektes und der Interaktion am Beispiel eines 2×2 Studiendesigns.

References

- Moshe, Isaac, Yannik Terhorst, Paula Philippi, Matthias Domhardt, Pim Cuijpers, Ioana Cristea, Laura Pulkki-Råback, Harald Baumeister, and Lasse B. Sander. 2021. "Digital Interventions for the Treatment of Depression: A Meta-Analytic Review." *Psychological Bulletin* 147 (8): 749–86. <https://doi.org/10.1037/bul0000334>.
- Wagner, Birgit, Andrea B. Horn, and Andreas Maercker. 2014. "Internet-Based Versus Face-to-Face Cognitive-Behavioral Intervention for Depression: A Randomized Controlled Non-Inferiority Trial." *Journal of Affective Disorders* 152–154 (January): 113–21. <https://doi.org/10.1016/j.jad.2013.06.032>.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(9) T-Tests

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Überblick

Kontinuum von ALM Designs

Extrem Szenario (1) Die Erwartungswerte aller Datenvariablen sind identisch.

$$y_i \sim N(\mu, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \Leftrightarrow \\ y = X\beta + \varepsilon, X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}, \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (1)$$

⇒ Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben.

Extrem Szenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden

$$y_i \sim N(\mu_i, \sigma^2) \text{ u.v. für } i = 1, \dots, n \Leftrightarrow \\ y = X\beta + \varepsilon \text{ mit } X := I_n \in \mathbb{R}^{n \times n}, \beta := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n, \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (2)$$

⇒ Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben.

⇒ Es gilt $\hat{\beta} = (I_n^T I_n)^{-1} I_n^T y = y$ und $\hat{\sigma}^2 = \frac{(y - I_n y)^T (y - I_n y)}{n - p} = 0$.

Beide Extrem Szenarien sind wissenschaftlich nicht ergiebig, da sie keine theoriegeleitete systematische Abhängigkeit zwischen der UV und der AV repräsentieren. Die im weiteren Verlauf der Vorlesung betrachteten ALM Designs liegen zwischen den beiden Extrem Szenarien und repräsentieren verschiedene Formen der systematischen Abhängigkeit zwischen UV und AV.

Faktorielle und Parametrische ALM Designs

Faktorielle ALM Designs

- Designmatrizen mit 1en und 0en, manchmal -1 en.
- Betaparameter repräsentieren Gruppenerwartungswerte.
- Betaparameterschätzer repräsentieren Gruppenstichprobenmittel.
- \Rightarrow T-Tests, Einfaktorielle Varianzanalyse, Mehrfaktorielle Varianzanalyse

Parametrische ALM Designs

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Regressor-Daten Kovarianzen.
- Es besteht ein enger Bezug zur Theorie der Korrelation.
- \Rightarrow Einfache lineare Regression, Multiple lineare Regression

Faktoriell-parametrische ALM Designs

- Designmatrizen mit mehreren faktoriellen und parametrischen Werten.
- Die parametrischen Regressoren werden oft als kontrollierte Kovariaten betrachtet.
- \Rightarrow Kovarianzanalyse

ALM Designs als Hypothesentestverfahren*

Testen von Unterschiedshypothesen

- T-Tests
- Einfaktorielle Varianzanalyse
- Mehrfaktorielle Varianzanalyse
- Kovarianzanalyse

Testen von Zusammenhangshypothesen

- Einfache lineare Regression/Korrelation
- Multiple lineare Regression/Multiple Korrelation

*Diese Sichtweise durch den Lehrenden nicht favorisiert.

T-Tests

Es gibt viele T-Test Varianten, jeweils mit eigenen Testgütefunktionen.

Wir fokussieren hier auf die Erkenntnis von T-Tests als Spezialfälle des ALMs.

Wir behandeln im Detail

- Einstichproben-T-Tests mit einfacher Nullhypothese und ungerichteter Alternativhypothese.
- Zweistichproben-T-Tests bei unabhängigen Stichproben unter Annahme identischer Varianzen mit einfacher Nullhypothese und ungerichteter Alternativhypothese.

Wir behandeln nicht

- T-Tests mit gerichteten Hypothesen oder einfachen Null- und Alternativhypothesen.
- Zweistichproben-T-Tests bei Annahme nicht identischer Varianzen (Behrens-Fischer Problem).
- Zweistichproben T-Tests bei abhängigen Stichproben.

Für dieses Themengebiet wird auf WTFI Einheiten (13) und (14) verwiesen.

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Einstichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Eine Gruppe (Stichprobe) randomisierter experimenteller Einheiten.

Annahme der unabhängigen und identischen Normalverteilung $N(\mu, \sigma^2)$ der Datenpunkte.

μ und σ^2 unbekannt.

Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ mit μ_0 beabsichtigt.

Anwendungsbeispiele

Pre-Post-Psychotherapie BDI Differenzanalyse einer Gruppe von Patient:innen

- $\mu \neq \mu_0 := 0 \Rightarrow$ Evidenz für Depressions symptomatikveränderung

Gruppenanalysen mit Wechsler Adult Intelligence Scale

- $\mu \neq \mu_0 := 100 \Rightarrow$ Evidenz für über- oder unterdurchschnittliche WAIS Performanz

Gruppenanalysen in der funktionellen Kernspintomographie

- $\mu > \mu_0 := 0 \Rightarrow$ Evidenz für regionale Gehirnaktivierung

Wir betrachten das Anwendungsbeispiel aus Einheit (8) Studiendesign und fokussieren auf die Gruppe (= Stichprobe, Experimentalbedingung) der Face-to-Face Therapie. Wir betrachten also den Datensatz der negativen PostBDI-PreBDI Differenzwerte mit Variablennamen "BDI".

Wir nehmen an, dass diese Datenpunkte u.i.v. Realisierungen von ZVen $y_i \sim N(\mu, \sigma^2)$ sind und nehmen weiter an, dass wir sind an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekanntem, Erwartungswertparameters μ im Sinne eines Hypothesentests interessiert sind.

Im Folgenden evaluieren diesen Datensatz zunächst im Sinne deskriptiver Statistiken, siehe dazu Einheit (11) Anwendungsbeispiel in Programmierung und Deskriptive Statistik.

Anwendungsszenario

Dateneinlesen

```
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv")  
D = read.table(fname, sep = ",", header = TRUE)
```

X	ID	Condition	PreBDI	PostBDI	BDI	Age	Duration
1	1	F2F	29	25	4	66	22
2	2	F2F	32	31	1	29	23
3	3	F2F	28	26	2	71	14
4	4	F2F	36	26	10	77	21
5	5	F2F	32	27	5	55	24
6	6	F2F	28	29	-1	50	24
7	7	F2F	33	27	6	31	16
8	8	F2F	33	27	6	20	17
9	9	F2F	33	25	8	73	23
10	10	F2F	30	26	4	28	13
11	11	F2F	36	27	9	21	23
12	12	F2F	32	25	7	76	16
13	13	F2F	29	29	0	38	18
14	14	F2F	24	22	2	30	18
15	15	F2F	35	28	7	44	21
16	16	F2F	31	22	9	48	17
17	17	F2F	31	26	5	46	14
18	18	F2F	34	25	9	51	16
19	19	F2F	34	25	9	71	20
20	20	F2F	33	27	6	23	13
21	21	F2F	34	21	13	53	22
22	22	F2F	33	31	2	61	15
23	23	F2F	31	22	9	59	18
24	24	F2F	25	26	-1	60	14
25	25	F2F	33	23	10	48	16
26	26	F2F	31	25	6	78	19
27	27	F2F	31	34	-3	44	15
28	28	F2F	26	27	-1	71	20
29	29	F2F	29	23	6	65	18
30	30	F2F	32	22	10	52	23
31	31	F2F	35	28	7	72	22
32	32	F2F	31	27	4	48	16
33	33	F2F	32	26	6	20	16
34	34	F2F	31	24	7	64	22
35	35	F2F	27	22	5	63	17
36	36	F2F	30	24	6	31	16
37	37	F2F	30	30	0	59	13
38	38	F2F	31	25	6	53	13
39	39	F2F	34	23	11	40	20
40	40	F2F	33	33	0	58	20

Anwendungsszenario

Histogramm

```
# Datensatz von Interesse
BDI_F2F = D$BDI[D$Condition == "F2F"] # BDI Differenzwerte in der F2F Gruppe

# Histogrammparameter
h = 1 # gewünschte Klassenbreite
b_0 = min(BDI_F2F) # b_0
b_k = max(BDI_F2F) # b_0
k = ceiling((b_k - b_0)/h) # Anzahl der Klassen
b = seq(b_0, b_k, by = h) # Klassen [b_{j-1}, b_j]
ylimits = c(0, 0.25) # y-Achsenlimits
xlimits = c(-5, 15) # x-Achsenlimits

# Abbildungsparameter
par( # für Details siehe ?par
     mfcol = c(1, 1), # 1 x 1 Panelstruktur
     family = "sans", # Serif-freier Fonttyp
     pty = "s", # Quadratische Abbildungsregion
     bty = "l", # L förmige Box
     las = 1, # Horizontale Achsenbeschriftung
     xaxs = "i", # x-Achse bei y = 0
     yaxs = "i", # y-Achse bei x = 0
     font.main = 1, # Non-Bold Titel
     cex = 1, # Textvergrößerungsfaktor
     cex.main = 1) # Titeltextvergrößerungsfaktor

# Histogramm
hist(
  BDI_F2F, # Delta.BDI Werte von Therapiebedingung i
  breaks = b, # Histogrammklassen
  freq = F, # normierte relative Häufigkeit
  xlim = xlimits, # x-Achsenlimits
  ylim = ylimits, # y-Achsenlimits
  xlab = "BDI", # x-Achsenbeschriftung
  ylab = "Geschätzte Wahrscheinlichkeit", # y-Achsenbeschriftung
  main = "") # Titelbeschriftung

# PDF Speicherung
dev.copy2pdf(
  file = file.path(getwd(), "9_Abbildungen", "alm_9_F2F_histogramm.pdf"),
  width = 4,
  height = 4)
```

Anwendungsszenario

Deskriptive Statistiken

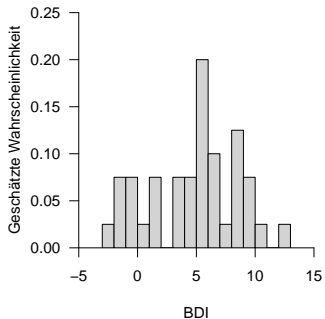
```
# Einlesen der Daten
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv")
D = read.table(fname, sep = ",", header = TRUE)

# Initialisierung eines Dataframes
tp = c("F2F") # Therapiebedingungen
ntp = length(tp) # Anzahl Therapiebedingungen
S = data.frame( # Dataframeerzeugung
  n = rep(NA, ntp), # Stichprobengrößen
  Max = rep(NA, ntp), # Maxima
  Min = rep(NA, ntp), # Minima
  Median = rep(NA, ntp), # Mediane
  Mean = rep(NA, ntp), # Mittelwerte
  Var = rep(NA, ntp), # Varianzen
  Std = rep(NA, ntp), # Standardabweichungen
  row.names = tp) # Therapiebedingungen

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data = D$BDI[D$Condition == tp[i]] # Daten
  S$n[i] = length(data) # Stichprobengröße
  S$Max[i] = max(data) # Maxima
  S$Min[i] = min(data) # Minima
  S$Median[i] = median(data) # Mediane
  S$Mean[i] = mean(data) # Mittelwerte
  S$Var[i] = var(data) # Varianzen
  S$Std[i] = sd(data) # Standardabweichungen
}
print(data)
```

```
> [1] 4 1 2 10 5 -1 6 6 8 4 9 7 0 2 7 9 5 9 9 6 13
> [22] 2 9 -1 10 6 -3 -1 6 10 7 4 6 7 5 6 0 6 11 0
```

Deskriptive Statistiken der negativen PostBDI-PreBDI Differenzen bei Face-to-Face Therapie



```
# Ausgabe  
print.AsIs(S)
```

```
>      n Max Min Median Mean  Var  Std  
> F2F 40 13  -3      6 5.28 14.8 3.85
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Definition (Einstichproben-T-Test Modell)

$y_i, i = 1, \dots, n$ seien Zufallsvariablen, die die n Datenpunkte eines Einstichproben-T-Test Anwendungsszenarios modellieren. Dann hat das *Einstichproben-T-Test Modell* die strukturelle Form

$$y_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (3)$$

die Datenverteilungsform

$$y_i \sim N(\mu, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (4)$$

und f\"ur den Datenvektor $y = (y_1, \dots, y_n)^T$ die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ und } \sigma^2 > 0. \quad (5)$$

Bemerkungen

- Das Modell ist identisch mit dem Modell unabhängiger und identisch normalverteilter Zufallsvariablen.
- Die Anzahl der Betaparameter ist $p = 1$.
- Die Äquivalenz der drei Modellformen wurde in Einheit (5) Modellformulierung ausführlich diskutiert.

Datensimulation (vgl. Einheit (5) Modellformulierung)

```
# Libraries
library(MASS) # Multivariate Normalverteilung

# Modellformulierung
n = 40 # Anzahl von Datenpunkten
p = 1 # Anzahl von Betaparameter
X = matrix(rep(1,n), nrow = n) # Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = 5 # wahrer, aber unbekannter, Betaparameter
sigsqr = 14 # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Theorem (Parameterschätzung im Einstichproben-T-Test Modell)

Gegeben sei die Designmatrixform des Einstichproben-T-Test Modells. Dann ergeben sich für den Betaparameterschätzer

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y}, \quad (6)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2 \quad (7)$$

Bemerkungen

- Die Formen von $\hat{\beta}$ und $\hat{\sigma}^2$ wurden in Einheit (6) Modellschätzung hergeleitet.
- \bar{y} und s_y^2 bezeichnen das Stichprobenmittel und die Stichprobenvarianz der y_1, \dots, y_n .

Modellschätzung

```
# Dateneinlesen
fname      = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D          = read.table(fname, sep = ",", header = TRUE)         # Dataframe
y          = D$BDI[D$Condition == "F2F"]                         # BDI Differenzwerte in der F2F Gruppe

# Modellformulierung
n          = length(y)                                          # Anzahl Datenpunkte
p          = 1                                                  # Anzahl Betaparameter
X          = matrix(rep(1,n), nrow = n)                        # Designmatrix

# Modellschätzung
beta_hat   = solve(t(X) %*% X) %*% t(X) %*% y                 # Betaparameterschätzer
eps_hat    = y - X %*% beta_hat                               # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)                 # Varianzparameterschätzer

# Ausgabe
cat("hat{beta}   : " , beta_hat,                               # Betaparameterschätzer
    "\nbar{y}    : " , mean(y),                               # Stichprobenmittel
    "\nhat{sigsqr} : " , sigsqr_hat,                           # Varianzparameterschätzer
    "\ns_y^2     : " , var(y))                                 # Stichprobenvarianz

> hat{beta}   : 5.28
> bar{y}      : 5.28
> hat{sigsqr} : 14.8
> s_y^2       : 14.8
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Überblick

- Wir gruppieren frequentistische Konfidenzintervalle und Hypothesentests unter Modellevaluation.
- Zu Konfidenzintervallen im Szenario u.i.n.v. ZVen siehe WTFI Einheit (11) Konfidenzintervalle.
- In der Praxis zielt die Evaluation von Einstichproben-T-Tests ALM Designs meist auf einen Hypothesentest.
- Die Theorie des Einstichproben-T-Tests ist umfangreich, siehe dazu WTFI (13) Einstichproben-T-Tests.
- Wir fokussieren hier auf zweiseitige Einstichproben-T-Tests mit ungerichteter Hypothese.
- Ein gutes Verständnis von WTFI Einheit (12) Hypothesentests wird im Folgenden vorausgesetzt.

Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Hypothesenszenarien

Im Folgenden näher betrachtetes Hypothesenszenario

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung (vgl. WTFI Einheiten (12) - (14))

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik
- (4) Test
- (5) Analyse der Testgütefunktion
- (6) Testumfangkontrolle
- (7) p-Werte
- (8) Analyse der Powerfunktion

(3) Teststatistik

Theorem (T-Teststatistik des Einstichproben-T-Tests)

Gegeben sei die Designmatrixform des Einstichproben-T-Test Modells. Dann ergibt sich für die T-Teststatistik mit

$$c := 1 \text{ und } c^T \beta_0 =: \mu_0, \quad (8)$$

dass

$$T = \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right) \quad (9)$$

und es gilt

$$T \sim t(\delta, n - 1) \text{ mit } \delta = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (10)$$

Bemerkungen

- Das Theorem basiert auf dem T-Teststatistik Theorem in Einheit (7) Modellevaluation.
- Die eng verwandte T-Statistik wurde bereits in Einheit (7) Modellevaluation hergeleitet.

(3) Teststatistik

Beweis

Mit dem T-Teststatistik Theorem in Einheit (7) Modellevaluation gilt

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y} - 1^T \mu_0}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right). \quad (11)$$

Weiterhin gilt mit demselben Theorem

$$\delta = \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} = \frac{1^T \mu - 1^T \mu_0}{\sqrt{\sigma^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (12)$$

(4) Test

Definition (Zweiseitiger Einstichproben-T-Tests)

Gegeben sei das Einstichproben-T-Test Modell. Für ein $\mu_0 \in \mathbb{R}$ seien die einfache Nullhypothese und die zusammengesetzte Alternativhypothese als

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (13)$$

definiert. Weiterhin sei die T-Teststatistik definiert als

$$T := \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right) \quad (14)$$

Dann ist der *zweiseitige Einstichproben-T-Tests* definiert als der kritische Wert-basierten Test

$$\phi(y) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (15)$$

Bemerkungen

- Ausführlicher handelt es sich um den *zweiseitigen Einstichproben-T-Tests mit ungerichteter Hypothese*.

(5) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testscenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; \delta, n - 1) + \psi(-k; \delta, n - 1) \quad (16)$$

wobei $\psi(\cdot; \delta, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

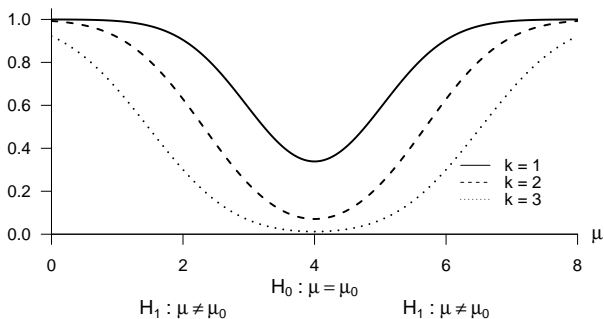
$$\delta := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (17)$$

und Freiheitsgradparameter $n - 1$ bezeichnet.

(5) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9$, $\mu_0 = 4$, $n = 12$ und $k = 1, 2, 3$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(5) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (18)$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt gleich sind, benötigen wir die also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{y} - \mu_0}{s_y} \right) \quad (19)$$

unter der Annahme $y_i \sim N(\mu, \sigma^2)$ u.i.v. für $i = 1, \dots, n$ nach einer nichtzentralen t -Verteilung $t(\delta, n - 1)$ mit Nichtzentralitätsparameter

$$\delta = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (20)$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen T-Tests ergibt sich, wie in ähnlicher Form bei der Betrachtung des zweiseitigen Z-Tests gesehen, zu

$$A =] - \infty, -k] \cup]k, \infty[. \quad (21)$$

(5) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k] \cup]k, \infty[) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k]) + \mathbb{P}_{\mu}(T \in [k, \infty[) \\&= \mathbb{P}_{\mu}(T \leq -k) + \mathbb{P}_{\mu}(T \geq k) \\&= \mathbb{P}_{\mu}(T \leq -k) + (1 - \mathbb{P}_{\mu}(T \leq k)) \\&= 1 - \mathbb{P}_{\mu}(T \leq k) + \mathbb{P}_{\mu}(T \leq -k) \\&= 1 - \psi(k; \delta, n - 1) + \psi(-k; \delta, n - 1),\end{aligned}\tag{22}$$

wobei $\psi(\cdot; \delta, n - 1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter $n - 1$ bezeichnet.

□

(6) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n - 1 \right), \quad (23)$$

wobei $\psi^{-1}(\cdot; n - 1)$ die inverse KVF der t -Verteilung mit $n - 1$ Freiheitsgraden ist.

(6) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k; \delta, n - 1) + \psi(-k; \delta, n - 1) \\ &= 1 - \psi(k; 0, n - 1) + \psi(-k; 0, n - 1) \\ &= 1 - \psi(k; n - 1) + \psi(-k; n - 1),\end{aligned}\tag{24}$$

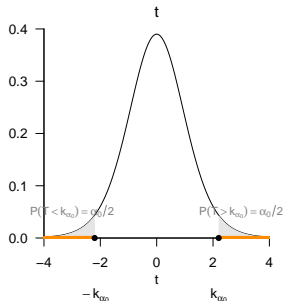
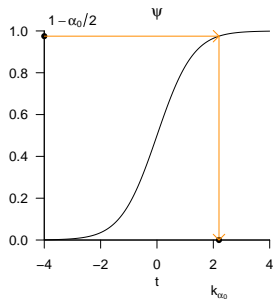
wobei $\psi(\cdot; \delta, n - 1)$ und $\psi(\cdot; n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter $n - 1$ sowie der t -Verteilung mit Freiheitsgradparameter $n - 1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}, n - 1) + \psi(-k_{\alpha_0}, n - 1) \\ &= 1 - \psi(k_{\alpha_0}, n - 1) + (1 - \psi(k_{\alpha_0}, n - 1)) \\ &= 2(1 - \psi(k_{\alpha_0}, n - 1)) \\ &= 2 \left(1 - \psi \left(\psi^{-1}(1 - \alpha_0/2, n - 1), n - 1 \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= \alpha_0,\end{aligned}\tag{25}$$

wobei die zweite Gleichung mit der Symmetrie der t -Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

(6) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



(6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein Datensatz v_1, \dots, v_n eine Realisation von $y_i \sim N(\mu, \sigma^2)$ u.i.v. für $i = 1, \dots, n$ mit unbekanntem Parameter μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von n, μ_0, \bar{v} und s_v berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{v} - \mu_0}{s_v} \right) \quad (26)$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(7) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (27)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n - 1)). \quad (28)$$

- Im Gegensatz zum Z-Test hängt bei T-Tests der p-Wert auch von der Stichprobengröße ab.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.076, für $T = 2.00$ und $n = 100$ ist der p-Wert dagegen 0.048.

(7) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \quad (29)$$

- Dies aber folgt aus

$$\begin{aligned} |t| &\geq \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq \psi\left(\psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(T \leq |t|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(T \geq |t|). \end{aligned} \quad (30)$$

Modellevaluation

Anwendungsszenario

```
# Dateneinlesen
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D = read.table(fname, sep = ",", header = TRUE) # Dataframe
y = D$BDI[D$Condition == "F2F"] # BDI Differenzwerte in der F2F Gruppe

# Modellformulierung
n = length(y) # Anzahl Datenpunkte
p = 1 # Anzahl Betaparameter
X = matrix(rep(1,n), nrow = n) # Designmatrix

# Modellschätzung
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer

# Modellevaluation
c = matrix(c(1), nrow = p) # Kontrastgewichtsvektor
mu_0 = 0 # Nullhypothese H_0
alpha_0 = 0.05 # Signifikanzniveau
k_alpha_0 = qt(1 - (alpha_0/2), n-1) # kritischer Wert
t_num = t(c) %*% beta_hat - mu_0 # T-Teststatistik Zähler
t_den = sqrt(sigsqr_hat %*% t(c)*solve(t(X) %*% X)%*%c) # T-Teststatistik Nenner
t = t_num/t_den # T-Teststatistik
if(abs(t) >= k_alpha_0){ # Test 1_{|T(X)| >= k_alpha_0}
  phi = 1 # Ablehnen von H_0
} else {
  phi = 0 # Nicht Ablehnen von H_0
}
pval = 2*(1 - pt(abs(t), n-1)) # p-Wert

> fg = 39
> t = 8.67
> alpha_0 = 0.05
> k_alpha_0 = 2.02
> phi = 1
> p-Wert = 1.25e-10
```

Anwendungsszenario

```
# Automatischer Einstichproben-T-Test
varphi = t.test(           # ?t.test für Details
  y,                       # Datensatz
  alternative = c("two.sided"), # H_1: \mu \neq \mu_0
  mu          = 0,         # \mu_0 (sic!)
  conf.level  = 1-alpha_0 # \delta = 1 - \alpha_0 (sic!)

# Ausgabe
print(varphi)
```

```
>
> One Sample t-test
>
> data: y
> t = 9, df = 39, p-value = 1e-10
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  4.04 6.51
> sample estimates:
> mean of x
>      5.28
```

```
# Genauere Ausgabe t
paste(varphi[1])
```

```
> [1] "c(t = 8.66623050649246)"
```

```
# Genauere Ausgabe p
paste(varphi[3])
```

```
> [1] "1.25177126014618e-10"
```

(8) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; \delta, n - 1) + \psi(-k_{\alpha_0}; \delta, n - 1) \quad (31)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

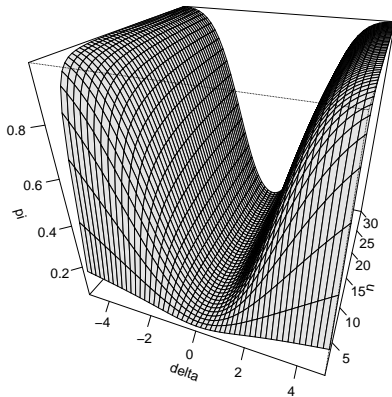
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n - 1) + \psi(-k_{\alpha_0}; \delta, n - 1) \quad (32)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert δ und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

(8) Analyse der Powerfunktion

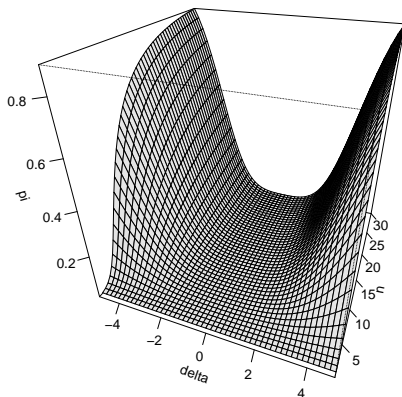
Powerfunktion für $\alpha_0 = 0.05$



Zweiseitige Einstichproben-T-Tests

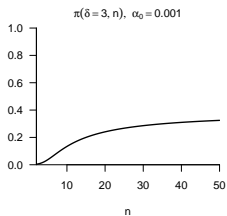
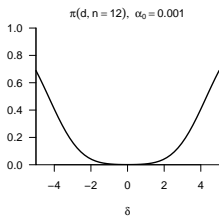
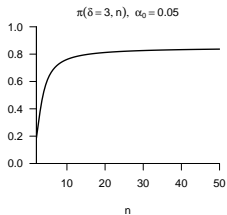
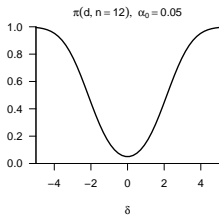
(8) Analyse der Powerfunktion

Powerfunktion für $\alpha_0 = 0.001$



(8) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



(8) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Wert $\delta = \sqrt{n} \frac{\mu - \mu_0}{\sigma}$ ab.

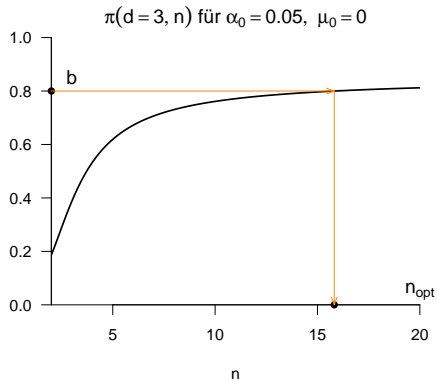
⇒ Wenn man δ schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert δ^* , den man mit $\pi(\delta, n) = b$ detektieren möchte.
- Ein konventioneller Wert ist $b = 0.8$.
- Man liest die für $\pi(\delta = \delta^*, n) = b$ nötige Stichprobengröße n ab.

(8) Analyse der Powerfunktion

Praktisches Vorgehen



Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

Zweistichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Zwei Gruppen (Stichproben) randomisierter experimenteller Einheiten.

Annahme unabhängiger identischer Normalverteilungen $N(\mu_1, \sigma^2)$ und $N(\mu_2, \sigma^2)$.

μ_1, μ_2 und σ^2 unbekannt.

Annahme eines identischen Varianzparameters für beide Gruppen.

Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

BDI Differenzwert Datenanalyse bei zwei Gruppen von Patient:innen

- Gruppe 1 Face-to-Face Therapie, Gruppe 2 Online Therapie
- $\mu_1 \neq \mu_2 \Leftrightarrow$ Unterscheiden sich die Therapiewirksamkeiten?

Forcierte Schwimmtestdatenanalyse bei zwei Gruppen genmanipulierter Mäuse

- Gruppe 1 Wildtyp, Gruppe 2 Serotoninrezeptormutation
- $\mu_1 \neq \mu_2 \Leftrightarrow$ Trägt Serotoninrezeptor zum Schwimmtestverhalten bei?

Wir betrachten das Anwendungsbeispiel aus Einheit (8) Studiendesign und fokussieren auf die Gruppen (= Stichproben, Experimentalbedingungen) der Face-to-Face und der Online Therapie. Wir betrachten dabei wieder den Datensatz der negativen PostBDI-PreBDI Differenzwerte mit Variablennamen "BDI".

Wir nehmen an, dass die Datenpunkte der Face-to-Face Therapiegruppe u.i.v. Realisierungen von ZVen $y_{1j} \sim N(\mu_1, \sigma^2)$ für $j = 1, \dots, 40$ und dass die Datenpunkte der Online Therapiegruppe u.i.v. Realisierungen von ZVen $y_{2j} \sim N(\mu_2, \sigma^2)$ für $j = 1, \dots, 40$ sind. Wir nehmen weiter an, dass wir an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich der wahren, aber unbekannt, Erwartungswertparameter μ_1 und μ_2 im Sinne eines Hypothesentests interessiert sind.

Im Folgenden evaluieren den hier betrachteten Datensatz zunächst im Sinne deskriptiver Statistiken, siehe dazu auch Einheit (11) Anwendungsbeispiel in Programmierung und Deskriptive Statistik.

Anwendungsszenario

Dateneinlesen | $j = 1, \dots, 20$ für jede Gruppe

```
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv")  
D = read.table(fname, sep = ",", header = TRUE)
```

	X	ID	Condition	PreBDI	PostBDI	BDI	Age	Duration
1	1	1	F2F	29	25	4	66	22
2	2	2	F2F	32	31	1	29	23
3	3	3	F2F	28	26	2	71	14
4	4	4	F2F	36	26	10	77	21
5	5	5	F2F	32	27	5	55	24
6	6	6	F2F	28	29	-1	50	24
7	7	7	F2F	33	27	6	31	16
8	8	8	F2F	33	27	6	20	17
9	9	9	F2F	33	25	8	73	23
10	10	10	F2F	30	26	4	28	13
11	11	11	F2F	36	27	9	21	23
12	12	12	F2F	32	25	7	76	16
13	13	13	F2F	29	29	0	38	18
14	14	14	F2F	24	22	2	30	18
15	15	15	F2F	35	28	7	44	21
16	16	16	F2F	31	22	9	48	17
17	17	17	F2F	31	26	5	46	14
18	18	18	F2F	34	25	9	51	16
19	19	19	F2F	34	25	9	71	20
20	20	20	F2F	33	27	6	23	13
41	41	41	ONL	31	27	4	70	23
42	42	42	ONL	31	25	6	63	16
43	43	43	ONL	34	29	5	41	23
44	44	44	ONL	34	29	5	28	14
45	45	45	ONL	30	24	6	43	22
46	46	46	ONL	30	33	-3	76	15
47	47	47	ONL	33	25	8	68	12
48	48	48	ONL	34	21	13	66	22
49	49	49	ONL	32	26	6	77	20
50	50	50	ONL	35	27	8	80	23
51	51	51	ONL	33	33	0	56	20
52	52	52	ONL	30	26	4	22	22
53	53	53	ONL	33	27	6	40	16
54	54	54	ONL	28	26	2	37	17
55	55	55	ONL	37	25	12	27	19
56	56	56	ONL	38	26	12	23	13
57	57	57	ONL	31	28	3	42	14
58	58	58	ONL	29	33	-4	40	19
59	59	59	ONL	34	29	5	30	21
60	60	60	ONL	32	30	2	57	22

Anwendungsszenario

Histogramme

```
# Histogrammparameter
h          = 1 # gewünschte Klassenbreite
b_0        = min(D$BDI) # b_0
b_k        = max(D$BDI) # b_0
k          = ceiling((b_k - b_0)/h) # Anzahl der Klassen
b          = seq(b_0, b_k, by = h) # Klassen [b_{j-1}, b_j]
ylimits    = c(0,.25) # y-Achsenlimits
xlimits    = c(-2,14) # x-Achsenlimits
therapie   = c("F2F", "ONL") # Therapiebedingungen
labs       = c("Face-to-Face", # Abbildungslabel
              "Online")

# Abbildungsparameter
par( # für Details siehe ?par
     mfcol = c(1,2), # 1 x 2 Panelstruktur
     family = "sans", # Serif-freier Fonttyp
     pty = "m", # Maximale Abbildungsregion
     bty = "l", # L förmige Box
     las = 1, # Horizontale Achsenbeschriftung
     xaxs = "i", # x-Achse bei y = 0
     yaxs = "i", # y-Achse bei x = 0
     font.main = 1, # Non-Bold Titel
     cex = 1, # Textvergrößerungsfaktor
     cex.main = 1) # Titeltextvergrößerungsfaktor

# Iteration über Therapiebedingungen
for(i in 1:2){
  hist(
    D$BDI[D$Condition == therapie[i]], # Werte von Therapiebedingung i
    breaks = b, # Histogrammklassen
    freq = F, # normierte relative Häufigkeit
    xlim = xlimits, # x-Achsenlimits
    ylim = ylimits, # y-Achsenlimits
    xlab = TeX("BDI"), # x-Achsenbeschriftung
    ylab = "", # y-Achsenbeschriftung
    main = labs[i]) # Titelbeschriftung
}

# PDF Speicherung
dev.copy2pdf(
  file = file.path(getwd(), "9_Abbildungen", "alm_9_F2F_ONL_histogramme.pdf"),
  width = 8,
  height = 4)
```

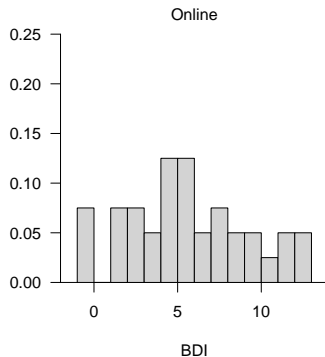
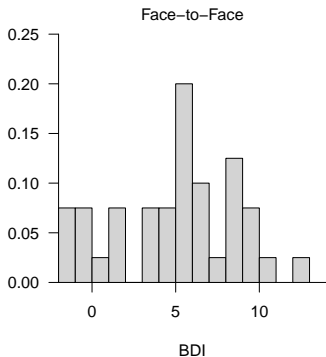
Deskriptive Statistiken

```
# Initialisierung eines Dataframes
ntp = c("F2F", "ONL")
n = length(tp)
S = data.frame(
  n = rep(NaN,ntp),
  Max = rep(NaN,ntp),
  Min = rep(NaN,ntp),
  Median = rep(NaN,ntp),
  Mean = rep(NaN,ntp),
  Var = rep(NaN,ntp),
  Std = rep(NaN,ntp),
  row.names = tp)

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data = D$BDI[D$Condition == tp[i]]
  S$n[i] = length(data)
  S$Max[i] = max(data)
  S$Min[i] = min(data)
  S$Median[i] = median(data)
  S$Mean[i] = mean(data)
  S$Var[i] = var(data)
  S$Std[i] = sd(data)
}
```

Anwendungsszenario

Deskriptive Statistiken der PostBDI-PreBDI Differenzen bei Face-to-Face und Online Therapie



```
# Ausgabe  
print.AsIs(S)
```

```
>      n Max Min Median Mean Var Std  
> F2F 40 13 -3   6 5.28 14.8 3.85  
> ONL 40 18 -5   6 5.92 25.7 5.07
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Definition (Zweistichproben-T-Test Modell)

y_{ij} mit $i = 1, 2$ und $j = 1, \dots, n_i$ seien Zufallsvariablen, die die Datenpunkte eines Zweistichproben-T-Test Anwendungsszenarios modellieren. Dann hat das *Zweistichproben-T-Test Modell* die strukturelle Form

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, 2, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (33)$$

die Datenverteilungsform

$$y_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. f\"ur } i = 1, 2, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (34)$$

und f\"ur den Datenvektor $y = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2})^T$ und $n := n_1 + n_2$ die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (35)$$

Bemerkungen

- i indiziert die Gruppen, j indiziert die Daten in jeder Gruppe.
- n_1 und n_2 repr\"asentieren die Gruppengr\"o\ss en, n repr\"asentiert die Gesamtanzahl an Datenpunkten.
- Die Anzahl der Betaparameter ist $p = 2$.
- Die \"Aquivalenz der drei Modellformen ergibt sich mit den Ergebnissen Einheit (5) Modellformulierung.

Datensimulation (vgl. Einheit (5) Modellformulierung)

```
# Libraries
library(MASS) # Multivariate Normalverteilung

# Modellformulierung
n_1 = 40 # Anzahl von Datenpunkten Gruppe 1
n_2 = 40 # Anzahl von Datenpunkten Gruppe 2
n = n_1 + n_2 # Gesamtanzahl Datenpunkte
p = 2 # Anzahl von Betaparameter
X = matrix(c(rep(1,n_1), rep(0,n_1), # Designmatrix
             rep(0,n_2), rep(1,n_2)),
           nrow = n)
I_n = diag(n) # n x n Einheitsmatrix
beta = matrix(c(1,2), nrow = p) # wahrer, aber unbekannter, Betaparameter
sigsqr = 14 # wahrer, aber unbekannter, Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
```


Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Theorem (Parameterschätzung im Zweistichproben-T-Test Modell)

Gegeben sei die Designmatrixform des Zweistichproben-T-Test Modells. Dann ergeben sich für den Betaparameterschätzer

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \quad (36)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \quad (37)$$

Bemerkungen

- \bar{y}_1 und \bar{y}_2 bezeichnen die gruppenspezifischen Stichprobenmittel.
- s_{12}^2 wird als *gepoolte Stichprobenvarianz* bezeichnet.
- Für einen Datensatz $y = (y_1, y_2)^T \in \mathbb{R}^{n_1+n_2}$ gilt allgemein, dass $s_y^2 \neq s_{12}^2$; die gepoolte Stichprobenvarianz und die Stichprobenvarianz eines konkatenierten Datensatzes sind im Allgemeinen also nicht identisch. Wir wollen das Konzept der gepoolten Stichprobenvarianz hier aber nicht weiter vertiefen.

Beweis

Für $i = 1, 2$ sei $y_i := (y_{i1}, \dots, y_{in_i})^T$. Dann ergibt sich für den Betaparameterschätzer

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left(\begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &= \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &=: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}.\end{aligned}\tag{38}$$

Beweis (fortgeführt)

Gleichsam ergibt sich für Varianzparameterschätzer mit $n = n_1 + n_2$ und $p = 2$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\ &= \frac{1}{n_1 + n_2 - 2} \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right)^T \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right) \\ &= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix} \\ &= \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \\ &=: s_{12}^2.\end{aligned}\tag{39}$$

Modellschätzung

```
# Dateneinlesen
fname      = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D          = read.table(fname, sep = ",", header = TRUE)         # Dataframe
y_1       = D$BDI[D$Condition == "F2F"]                          # BDI Differenzwerte in der F2F Gruppe
y_2       = D$BDI[D$Condition == "ONL"]                          # BDI Differenzwerte in der ONL Gruppe

# Modellformulierung
n_1       = length(y_1)                                          # Anzahl Datenpunkte Gruppe 1 (F2F)
n_2       = length(y_2)                                          # Anzahl Datenpunkte Gruppe 2 (ONL)
n         = n_1 + n_2                                           # Gesamtanzahl Datenpunkte
y         = matrix(c(y_1, y_2), nrow = n)                        # Datenvektor
p         = 2                                                    # Anzahl Betaparameter
X         = matrix(c(rep(1,n_1), rep(0,n_1),                    # Designmatrix
                    rep(0,n_2), rep(1,n_2)),
                  nrow = n)

# Modellschätzung
beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y                    # Betaparameterschätzer
eps_hat   = y - X %*% beta_hat                                  # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)                   # Varianzparameterschätzer
s_sqr_12  = ((n_1-1)*var(y_1) + (n_2-1)*var(y_2)) / (n_1+n_2-2) # gepoolte Stichprobenvarianz

# Ausgabe
cat("hat{beta}      : ", paste(beta_hat),                        # Betaparameterschätzer
    "\nbar{y}_1, bar{y}_2 : ", paste( c(mean(y_1), mean(y_2))), # Stichprobenmittel
    "\nhat{sigsqr}     : ", paste(sigsqr_hat),                  # Varianzparameterschätzer
    "\ns_12^2         : ", paste(s_sqr_12),                      # gepoolte Stichprobenvarianz
    "\ns_y^2          : ", paste(var(y)))                         # Stichprobenvarianz des konkatentierten Datensatzes

> hat{beta}      : 5.275 5.925
> bar{y}_1, bar{y}_2 : 5.275 5.925
> hat{sigsqr}    : 20.2403846153846
> s_12^2        : 20.2403846153846
> s_y^2         : 20.0911392405063
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Überblick

- Wir gruppieren frequentistische Konfidenzintervalle und Hypothesentests unter Modellevaluation.
- Wir verzichten an dieser Stelle auf eine Diskussion von Konfidenzintervallen
- In der Praxis zielt die Evaluation von Zweistichproben-T-Tests ALM Designs meist auf einen Hypothesentest.
- Die Theorie der Zweistichproben-T-Tests ist umfangreich, siehe dazu WTFI (14) Zweistichproben-T-Tests.
- Ein gutes Verständnis von WTFI Einheit (12) Hypothesentests wird im Folgenden vorausgesetzt.

Im Zweistichproben-T-Test ALM Design ergeben sich folgende Hypothesenszenarien

- $H_0 : \mu_1 - \mu_2 = \mu_0$ und $H_1 : \mu_1 - \mu_2 \neq \mu_0$
- $H_0 : \mu_1 - \mu_2 \leq \mu_0$ und $H_1 : \mu_1 - \mu_2 > \mu_0$
- $H_0 : \mu_1 - \mu_2 \geq \mu_0$ und $H_1 : \mu_1 - \mu_2 < \mu_0$

Wir betrachten hier nur exemplarisch $H_0 : \mu_1 - \mu_2 = \mu_0$ und $H_1 : \mu_1 - \mu_2 \neq \mu_0$.

Für $\mu_0 := 0$ gelten dabei insbesondere

- $H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow H_0 : \mu_1 = \mu_2$
- $H_1 : \mu_1 - \mu_2 \neq 0 \Leftrightarrow H_1 : \mu_1 \neq \mu_2$

Gliederung (vgl. WTFI Einheiten (12) - (14))

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik
- (4) Test
- (5) Analyse der Testgütefunktion
- (6) Testumfangkontrolle
- (7) p-Werte
- (8) Analyse der Powerfunktion

(3) Teststatistik

Theorem (T-Teststatistik des Zweistichproben-T-Tests)

Gegeben sei die Designmatrixform des Zweistichproben-T-Tests. Dann ergibt sich für die T-Teststatistik mit

$$c := (1, -1)^T \text{ und } c^T \beta_0 =: \mu_0, \quad (40)$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (41)$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (42)$$

Bemerkungen

- Das Theorem basiert auf dem T-Teststatistik Theorem in Einheit (7) Modellevaluation.

(3) Teststatistik

Beweis

Mit dem T-Teststatistik Theorem in Einheit (7) Modellevaluation gilt zunächst für die Zähler von T und δ , dass

$$c^T \hat{\beta} - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} - \mu_0 = \bar{y}_1 - \bar{y}_2 - \mu_0 \quad (43)$$

und

$$c^T \beta - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0, \quad (44)$$

respektive. Weiterhin gilt für die Nenner von T und δ , dass

$$c^T (X^T X)^{-1} c = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} n_1^{-1} & -n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2} \quad (45)$$

Außerdem gilt

$$\left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_1 + n_2}{n_1 n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_1 n_2}{n_1 + n_2} \right)^{\frac{1}{2}} \quad (46)$$

Zusammengenommen folgt direkt, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad \text{und} \quad \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (47)$$

(4) Test

Definition (Zweiseitiger Zweistichproben-T-Test)

Gegeben sei das Zweistichproben-T-Test Modell. Für ein $\mu_0 \in \mathbb{R}$ seien die einfache Nullhypothese und die zusammengesetzte Alternativhypothese gegeben durch

$$H_0 : \mu_1 - \mu_2 = \mu_0 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 - \mu_2 = \mu_0\} \quad (48)$$

und

$$H_1 : \mu_1 - \mu_2 \neq \mu_0 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 - \mu_2 \neq \mu_0\}, \quad (49)$$

respektive. Weiterhin sei die T-Teststatistik definiert durch

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{y}_1 - \bar{y}_2}{s_{12}} \right) \quad (50)$$

Dann ist der zweiseitige *Zweistichproben-T-Teststatistik* definiert als der kritischen Wert-basierte Test

$$\phi(y) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (51)$$

Bemerkungen

- Ausführlicher handelt es sich um den *zweiseitigen Zweistichproben-T-Test mit ungerichteter Hypothese*.

(5) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

Es sei ϕ der im obigen Modell formulierte Zweistichproben-T-Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ := 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2) \quad (52)$$

wobei $\psi(\cdot; \delta, n_1 + n_2 - 2)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$\delta := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (53)$$

und Freiheitsgradparameter $n_1 + n_2 - 2$ bezeichnet.

Bemerkungen

- q_ϕ ist eine bivariate reellwertige Funktion.
- q_ϕ kann alternativ als univariate reellwertige Funktion von $\Delta := \mu_1 - \mu_2$ konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

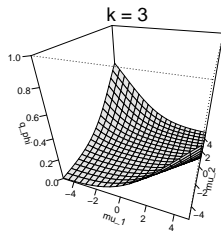
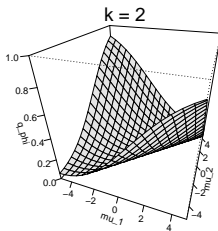
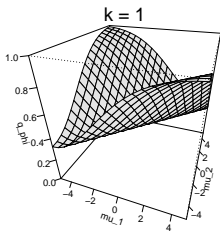
$$n \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 \quad (54)$$

- Für einen Beweisansatz, siehe DeGroot and Schervish (2012) Seite 591.

(5) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9, n_1 = 12, n_2 = 12$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(6) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (55)$$

wobei $\psi^{-1}(\cdot; n_1 + n_2 - 2)$ die inverse KVF der t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden ist.

Bemerkungen

- Das Resultat folgt in Analogie zum Einstichproben-T-Test.
- Im Vergleich zum Einstichproben-T-Testfall gilt lediglich

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (56)$$

(6) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass die Daten zweier Gruppen v_{11}, \dots, v_{1n_1} und v_{21}, \dots, v_{2n_2} Realisationen von $y_{1j} \sim N(\mu_1, \sigma^2)$ u.i.v. für $j = 1, \dots, n_1$ und $y_{2j} \sim N(\mu_2, \sigma^2)$ u.i.v. für $j = 1, \dots, n_2$ mit unbekanntem Parametern μ_1, μ_2, σ^2 sind.
- Man möchte entscheiden, ob eher $H_0 : \mu_1 - \mu_2 = \mu_0$ oder $H_1 : \mu_1 - \mu_2 \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n_1 = 12, n_2 = 12$, also Freiheitsgradparameter $12 + 12 - 2 = 22$, so dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 22) \approx 2.07$ ist.
- Anhand von $n_1, n_2, \bar{v}_1, \bar{v}_2$ und der gepoolten Stichprobenstandardabweichung s_{12} berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{v}_1 - \bar{v}_2}{s_{12}} \right) \quad (57)$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Zweistichproben-T-Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(7) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n_1 + n_2 - 2)$ abgelehnt werden. Für diese α_0 gilt, wie bereits mehrfach gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (58)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (59)$$

- Im Vergleich zum Einstichprobenfall gilt lediglich $n \leftrightarrow n_1 + n_2 - 2$.

Modellevaluation

Anwendungsszenario

```
# Dateneinlesen
fname = file.path(getwd(), "9_Daten", "data_9_t_tests.csv") # Dateiname
D = read.table(fname, sep = ",", header = TRUE) # Dataframe
y_1 = D$BDI[D$Condition == "F2F"] # BDI Differenzwerte in der F2F Gruppe
y_2 = D$BDI[D$Condition == "ONL"] # BDI Differenzwerte in der ONL Gruppe

# Modellformulierung
n_1 = length(y_1) # Anzahl Datenpunkte Gruppe 1 (F2F)
n_2 = length(y_2) # Anzahl Datenpunkte Gruppe 2 (ONL)
n = n_1 + n_2 # Gesamtanzahl Datenpunkte
y = matrix(c(y_1, y_2), nrow = n) # Datenvektor
p = 2 # Anzahl Betaparameter
X = matrix(c(rep(1, n_1), rep(0, n_1), # Designmatrix
             rep(0, n_2), rep(1, n_2)),
           nrow = n)

# Modellschätzung
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer

# Modellevaluation
c = matrix(c(1, -1), nrow = 2) # Kontrastgewichtsvektor
mu_0 = 0 # Nullhypothese H_0
alpha_0 = 0.05 # Signifikanzniveau
k_alpha_0 = qt(1 - (alpha_0/2), n-1) # kritischer Wert
t_num = t(c) %*% beta_hat - mu_0 # T-Teststatistik Zähler
t_den = sqrt(sigsqr_hat*t(c) %*% solve(t(X) %*% X)%*%c) # T-Teststatistik Nenner
t = t_num/t_den # T-Teststatistik
if(abs(t) >= k_alpha_0){ # Test  $1_{\{|T(X)| \geq k_{\alpha_0}\}}$ 
  phi = 1 # Ablehnen von  $H_0$ 
} else {
  phi = 0 # Nicht Ablehnen von  $H_0$ 
}
pval = 2*(1-pt(abs(t), n_1+n_2-2)) # p-Wert

> fg = 78
> t = -0.646
> alpha_0 = 0.05
> k_alpha_0 = 1.99
> phi = 0
> p-Wert = 0.52
```

Anwendungsszenario

```
# Automatischer Zweistichproben-T-Test
varphi = t.test(
  y_1,
  y_2,
  var.equal = TRUE,
  alternative = c("two.sided"),
  conf.level = 1-alpha_0)

# ?t.test für Details
# Datensatz y_1
# Datensatz y_2
# \sigma_1^2 = \sigma_2^2
# H_1: \mu_1 \neq \mu_2
# \delta = 1 - \alpha_0 (sic!)
```

```
# Ausgabe
print(varphi)
```

```
>
> Two Sample t-test
>
> data: y_1 and y_2
> t = -0.6, df = 78, p-value = 0.5
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
> -2.65 1.35
> sample estimates:
> mean of x mean of y
> 5.28 5.92
```

```
# Genauere Ausgabe t
paste(varphi[1])
```

```
> [1] "c(t = -0.646128613569434)"
```

```
# Genauere Ausgabe p
paste(varphi[3])
```

```
> [1] "0.520092604281206"
```

(8) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ := 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2) \quad (60)$$

als Funktion des Nichtzentralitätsparameters $\delta := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma}$ und der Summe der Stichprobenumfänge $n := n_1 + n_2$ bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 2)$ mit festem α_0 . Es ergibt sich die multivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n - 2) + \psi(-k_{\alpha_0}; \delta, n - 2) \quad (61)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert δ und von der Summe der Stichprobengrößen n ab. De-facto handelt es sich also um die gleiche Powerfunktion wie beim zweiseitigen Einstichproben-T-Test mit dem einzigen Unterschied, dass für den Freiheitsgradparameter $n - 2$ anstelle von $n - 1$ gilt. Wir verzichten auf eine erneute Visualisierung.

(8) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem $n = n_1 + n_2$ steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.
- Ungleichgewichte zwischen n_1 und n_2 werden durch die Tatsache ausglich, dass Datenpunkte einer Stichproben auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Wert $\delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma}$ ab.

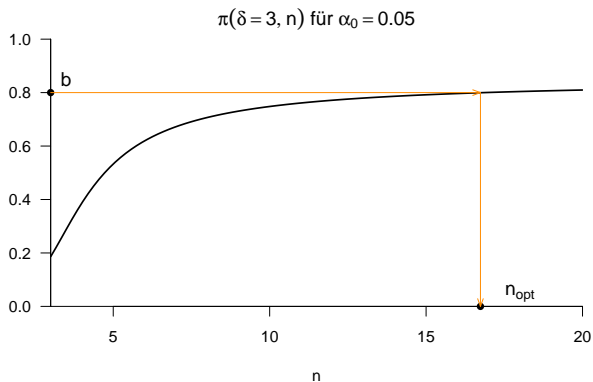
⇒ Wenn man δ schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert δ^* , den man mit $\pi(\delta, n) = b$ detektieren möchte.
- Ein konventioneller Wert ist $b = 0.8$.
- Man liest die für $\pi(\delta = \delta^*, n) = b$ nötige Stichprobengröße n ab.

(8) Analyse der Powerfunktion

Praktisches Vorgehen



Selbstkontrollfragen

1. Erläutern Sie die Extremszenarien im Kontinuum von ALM Designs.
2. Erläutern Sie die Begriffe der faktoriellen und parametrischen ALM Designs.
3. Nennen Sie Beispiele für faktorielle, parametrische, und faktoriell-parametrische ALM Designs.
4. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
5. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
6. Geben Sie die Definition des Einstichproben-T-Test Modells wieder.
7. Geben Sie das Theorem zur Parameterschätzung im Einstichproben-T-Test Modell wieder.
8. Geben Sie das Theorem zur T-Teststatistik des Einstichproben-T-Tests wieder.
9. Geben Sie die Definition des zweiseitigen Einstichproben-T-Tests (mit ungerichteter Hypothese) wieder.
10. Skizzieren Sie die Testgütefunktion des zweiseitigen Einstichproben-T-Tests.
11. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level- α_0 -Einstichproben-T-Test wieder.
12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Einstichproben-T-Tests.
13. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Einstichproben-T-Test wieder.
14. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Einstichproben-T-Tests ab?
15. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.
16. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.
17. Betrachten Sie die PostBDI-PreBDI Differenzwertdaten der Waitlist Control Bedingung im Beispieldatensatz. Erstellen Sie ein Histogramm dieser Daten und evaluieren Sie Ihnen bekannte deskriptive Statistiken zu diesen Daten. Führen Sie einen zweiseitigen Einstichproben-T-Test mit Nullhypothesenparameter $\mu_0 = 0$ durch. Dokumentieren Sie Ihre Ergebnisse. Was folgern Sie aus den sich ergebenden Resultaten?

Selbstkontrollfragen

18. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.
19. Geben Sie die Definition des Zweistichproben-T-Test Modells wieder.
20. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test Modell wieder.
21. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.
22. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.
23. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests (mit ungerichteter Hypothese) wieder.
24. Skizzieren Sie die Testgütefunktion des zweiseitigen Zweistichproben-T-Tests.
25. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level- α_0 -Zweistichproben-T-Test wieder.
26. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Zweistichproben-T-Tests.
27. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Zweistichproben-T-Test wieder.
28. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Zweistichproben-T-Tests ab?
29. Skizzieren Sie die Powerfunktion des zweiseitigen Zweistichproben-T-Tests bei fester Stichprobengröße.
30. Skizzieren Sie die Powerfunktion des zweiseitigen Zweistichproben-T-Tests bei festem Nichtzentralitätsparameter.
31. Betrachten Sie die Daten zum Alter der Patient:innen in der Face-to-Face und Online Therapie Bedingung im Beispieldatensatz. Erstellen gruppenspezifische Histogramme dieser Daten und evaluieren Sie gruppenspezifische deskriptive Statistiken zu diesen Daten. Führen Sie einen zweiseitigen Zweistichproben-T-Test mit Nullhypothesenparameter $\mu_0 = 0$ durch. Dokumentieren Sie Ihre Ergebnisse. Was folgern Sie aus den sich ergebenden Resultaten?

References

DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(10) Einfaktorielle Varianzanalyse

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Anwendungsszenario

Zwei oder mehr Gruppen (Stichproben) randomisierter experimenteller Einheiten.

Annahme der unabhängigen und identischen Normalverteilung $N(\mu_i, \sigma^2)$ der Daten.

μ_i und σ^2 unbekannt.

Absicht des inferentiellen Testens der Nullhypothese identischer Gruppenerwartungswerte.

⇒ Generalisierung des Zweistichproben-T-Tests bei unabhängigen Stichproben mit einfacher Nullhypothese für mehr als Zweistichproben

Anwendungsbeispiele

PrePost-BDI Differenzwertanalyse für drei Gruppen von Patient:innen (F2F, ONL, WLC)

- Inferentielle Evidenz für Gruppenerwartungswertunterschiede?
- Evidenz für unterschiedliche Therapiewirksamkeit?

Tense Arousal Differenzwertanalyse bei vier Filmclip-basierten Emotionsinduktionsbedingungen

- Inferentielle Evidenz für Gruppenerwartungswertunterschiede?
- Evidenz für unterschiedliche Tense Arousal Induktion?

Anwendungsszenario

Dateneinlesen

```
fname = file.path(getwd(), "10_Daten", "10_Einfaktorielle_Varianzanalyse_Daten.csv")  
D = read.table(fname, sep = ",", header = TRUE)
```

	ID	Condition	PreBDI	PostBDI	BDI	Age	Duration
1	1	F2F	29	25	4	66	22
2	2	F2F	32	31	1	29	23
3	3	F2F	28	26	2	71	14
4	4	F2F	36	26	10	77	21
5	5	F2F	32	27	5	55	24
6	6	F2F	28	29	-1	50	24
7	7	F2F	33	27	6	31	16
8	8	F2F	33	27	6	20	17
9	9	F2F	33	25	8	73	23
10	10	F2F	30	26	4	28	13
41	41	ONL	31	27	4	70	23
42	42	ONL	31	25	6	63	16
43	43	ONL	34	29	5	41	23
44	44	ONL	34	29	5	28	14
45	45	ONL	30	24	6	43	22
46	46	ONL	30	33	-3	76	15
47	47	ONL	33	25	8	68	12
48	48	ONL	34	21	13	66	22
49	49	ONL	32	26	6	77	20
50	50	ONL	35	27	8	80	23
81	81	WLC	27	31	-4	59	21
82	82	WLC	29	35	-6	28	21
83	83	WLC	33	35	-2	41	14
84	84	WLC	24	29	-5	64	18
85	85	WLC	31	23	8	74	19
86	86	WLC	30	38	-8	62	18
87	87	WLC	32	32	0	34	17
88	88	WLC	28	32	-4	59	17
89	89	WLC	30	30	0	37	12
90	90	WLC	30	32	-2	77	23

Boxplot

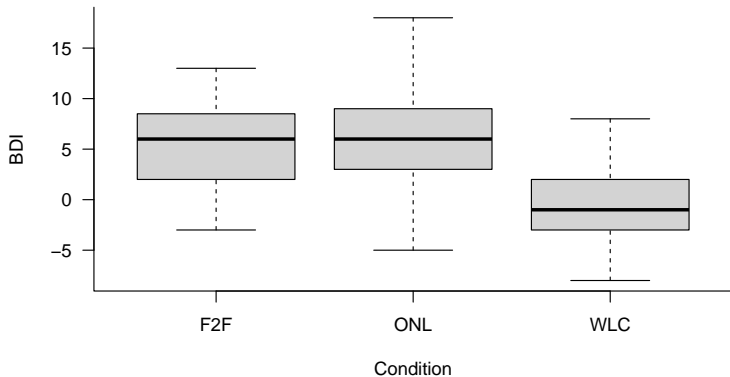
```
# Dateneinlesen
fname      = file.path(getwd(), "10_Daten", "10_Einfaktorielle_Varianzanalyse_Daten.csv")
D          = read.table(fname, sep = ",", header = TRUE)

# Abbildungsparameter
par(
family     = "sans",           # für Details siehe ?par
pty        = "m",             # Serif-freier Fonttyp
bty        = "l",             # Maximale Abbildungsregion
lwd        = 1,               # L förmige Box
las        = 1,               # Liniendicke
font.main  = 1,               # Horizontale Achsenbeschriftung
cex        = 1,               # Non-Bold Titel
cex.main   = 1.2,            # Textvergrößerungsfaktor
                                # Titeltextrvergrößerungsfaktor

# Boxplot
boxplot(BDI ~ Condition, data = D)      # BDI ~ Condition enkodiert die Datengruppierung

# PDF Speicherung
dev.copy2pdf(
file       = file.path(getwd(), "10_Abbildungen", "alm_10_aov_1_boxplot.pdf"),
width      = 7,
height     = 4)
```

Boxplot



Anwendungsszenario

Balkendiagramm

```
# Datenselektion
fname = file.path(getwd(), "10_Daten", "10_Einfaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# Datenselektion
A = data.frame(
  F2F = D$BDI[D$Condition == "F2F"], # F2F BDI Daten
  ONL = D$BDI[D$Condition == "ONL"], # ONL BDI Daten
  WLC = D$BDI[D$Condition == "WLC"]) # WLC BDI Daten

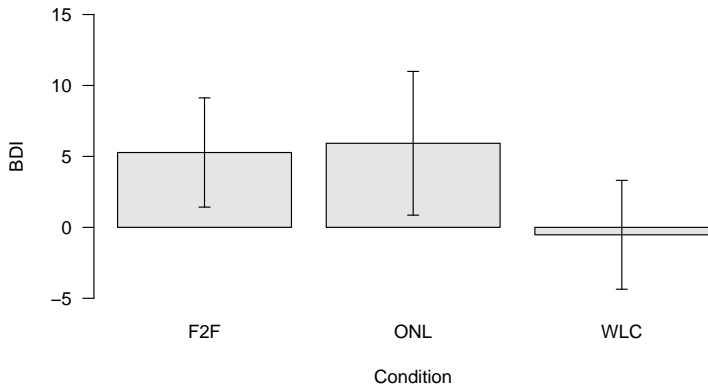
# Deskriptive Statistiken
groupmeans = colMeans(A) # Gruppenmittelwerte
groupstds = apply(A,2,sd) # Gruppenstandardabweichungen

# Balkendiagramm
par(
  family = "sans", # für Details siehe ?par
  pty = "m", # Serif-freier Fonttyp
  bty = "n", # Maximale Abbildungsregion
  lwd = 1, # L förmige Box
  las = 1) # Liniendicke
x = barplot(
  groupmeans, # Horizontale Achsenbeschriftung
  ylim = c(-5,15), # Ausgabe der x-Ordinaten (?barplot für Details)
  col = "gray90",
  ylab = "BDI",
  xlab = "Condition")
arrows(
  x0 = x, # für Details siehe ?arrows
  y0 = groupmeans - groupstds, # arrow start x-ordinate
  x1 = x, # arrow start y-ordinate
  y1 = groupmeans + groupstds, # arrow end x-ordinate
  code = 3, # arrow end y-ordinate
  angle = 90, # Pfeilspitzen beiderseits
  length = 0.05) # Pfeilspitzenwinkel -> Linie
# Linielänge

# PDF Speicherung
dev.copy2pdf(
  file = file.path(getwd(), "10_Abbildungen", "alm_10_aov_1_barplot.pdf"),
  width = 7,
  height = 4)
```

Balkendiagramm

Gruppenmittelwerte \pm Gruppenstandardabweichungen



Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Definition (EVA Modell in Erwartungswertparameterdarstellung)

y_{ij} mit $i = 1, \dots, p$, welche die Gruppen indizieren, und $j = 1, \dots, n_i$, welche die experimentellen innerhalb der Gruppen indizieren, seien Zufallsvariablen, die die Datenpunkte eines Einfaktorien Varianzanalyse (EVA) Szenarios modellieren. Dann hat das *EVA Modell in Erwartungswertparameterdarstellung* die strukturelle Form

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, p, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R}, \sigma^2 > 0. \quad (1)$$

und die Datenverteilungsform

$$y_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, p, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R}, \sigma^2 > 0. \quad (2)$$

Wenn $n_i := m$ für alle $i = 1, \dots, p$ heißt das EVA Szenario *balanciert*.

Bemerkungen

- Die Äquivalenz der beiden Modellformulierungen folgt aus den Ergebnissen in Einheit (5) Modellformulierung
- Es handelt sich um die Generalisierung des Zweistichproben-T-Test Modells für unabhängige Stichproben unter Annahme identischer Varianzparameter von $p = 2$ auf ein beliebiges $p \in \mathbb{N}$.
- Bei balancierten Varianzanalyseszenarien besteht jede Datengruppe aus der gleichen Anzahl von Datenpunkten.

Motivation der Effektdarstellung

Die Erwartungswertparameterdarstellung des EVA Modells ist ein valides ALM, das sich in dieser Form auch in der Literatur findet (z.B. Georgii (2009), Kapitel 12.4). Im Sinne der Konsistenz mit mehrfaktoriellen Varianzanalysemodellen bietet sich jedoch eine Reparameterisierung der Erwartungswertparameter an. Kern dieser Reparameterisierung ist es, den Erwartungswertparameter der i ten Gruppe als Summe eines *gruppenübergreifenden Erwartungswertparameters* $\mu_0 \in \mathbb{R}$ und eines *gruppenspezifischen Effektparameters* $\alpha_i \in \mathbb{R}$ zu modellieren,

$$\mu_i := \mu_0 + \alpha_i \text{ für } i = 1, \dots, p. \quad (3)$$

Dabei modelliert α_i den Unterschied (die Differenz) zwischen dem i ten Erwartungswertparameter μ_i und dem gruppenübergreifenden Erwartungswertparameter μ_0 ,

$$\alpha_i = \mu_i - \mu_0 \text{ für } i = 1, \dots, p. \quad (4)$$

Allerdings hat die in dieser Form vorgenommene Reparameterisierung einen entscheidenden Nachteil: es werden p Erwartungswertparameter $\mu_i, i = 1, \dots, p$ durch die $p + 1$ Parameter μ_0 und $\alpha_i, i = 1, \dots, p$ dargestellt. Diese Darstellung ist im Allgemeinen nicht eindeutig. Zum Beispiel können die Erwartungswertparameter $\mu_1 = 3, \mu_2 = 5, \mu_3 = 6$ sowohl durch den gruppenspezifischen Erwartungswertparameter $\mu_0 = 0$ und die gruppenunspezifischen Effektparameter $\alpha_1 = 3, \alpha_2 = 5, \alpha_3 = 6$ als auch durch den gruppenunspezifischen Erwartungswertparameter $\mu_0 = 1$ und die gruppenspezifischen Effektparameter $\alpha_1 = 2, \alpha_2 = 4, \alpha_3 = 5$ dargestellt werden. Man sagt in diesem Kontext auch, dass das EVA Modell mit (3) *überparameterisiert* ist.

Motivation der Effektdarstellung

Datenanalytisch hat die Überparameterisierung eines Varianzanalysemodells den Nachteil, dass aus p geschätzten Erwartungswertparametern $p + 1$ Betaparameterschätzer bestimmt werden müssten, was wie oben gesehen nicht eindeutig erfolgen kann. Um diese Probleme in der Effektparameterdarstellung des EVA Modells zu umgehen und diese konsistent auf mehrfaktorielle Varianzanalysemodelle zu übertragen, bietet sich die Einführung der Nebenbedingung

$$\alpha_1 := 0 \tag{5}$$

an. Es wird also ein Effektparameter von vornherein als "gleich Null" angenommen. Für die gruppenspezifischen Erwartungswertparameter ergibt sich damit

$$\begin{aligned} \mu_1 &:= \mu_0 \\ \mu_i &:= \mu_0 + \alpha_i \text{ für } i = 2, \dots, p. \end{aligned} \tag{6}$$

Hierbei wird die erste Gruppe nun als *Referenzgruppe* bezeichnet und die α_i modellieren die Differenz zwischen dem Erwartungswertparameter der i ten Gruppe und dem Erwartungswertparameter der ersten Gruppe:

$$\alpha_i = \mu_i - \mu_0 = \mu_i - \mu_1 \text{ für } i = 1, \dots, p. \tag{7}$$

μ_0 ist also kein gruppenübergreifender Erwartungswertparameter mehr, sondern identisch mit dem Erwartungswertparameter der ersten Gruppe. Welche tatsächliche experimentelle Gruppe dabei als "erste Gruppe" definiert wird, ist unerheblich. Entscheidend ist, dass die entsprechenden Erwartungswertparameterschätzer $\hat{\mu}_0, \hat{\alpha}_2, \dots, \hat{\alpha}_p$ korrekt als (1) Erwartungswertparameterschätzer der Referenzgruppe ($\hat{\mu}_0$) und (2) geschätzte Erwartungswertparameterdifferenz zwischen dem Erwartungswertparameter der Referenzgruppe und der dem Erwartungswertparameter der i ten Gruppe ($\hat{\alpha}_2, \dots, \hat{\alpha}_p$) verstanden werden. Wir formalisieren das oben Gesagte in folgendem Theorem.

Theorem (EVA Modell in Effektdarstellung mit Referenzgruppe I)

Gegeben sei das EVA Modell in Erwartungswertparameterdarstellung. Dann können die Zufallsvariablen, die die Datenpunkte des EVA Szenarios modellieren, äquivalent in der strukturellen Form

$$\begin{aligned}y_{1j} &= \mu_0 + \varepsilon_{1j} && \text{mit } \varepsilon_{1j} \sim N(0, \sigma^2) \text{ u.i.v. für } j = 1, \dots, n_1 \\y_{ij} &= \mu_0 + \alpha_i + \varepsilon_{ij} && \text{mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 2, \dots, p, j = 1, \dots, n_i\end{aligned} \quad (8)$$

mit

$$\alpha_i := \mu_i - \mu_1 \text{ für } i = 2, \dots, p \quad (9)$$

und in der entsprechenden Datenverteilungsform

$$\begin{aligned}y_{1j} &\sim N(\mu_0, \sigma^2) && \text{u.i.v. für } j = 1, \dots, n_1 \text{ mit } \mu_1 \in \mathbb{R}, \sigma^2 > 0 \\y_{ij} &\sim N(\mu_0 + \alpha_i, \sigma^2) && \text{u.i.v. für } i = 2, \dots, p, j = 1, \dots, n_i \text{ mit } \alpha_i \in \mathbb{R}, \sigma^2 > 0\end{aligned} \quad (10)$$

geschrieben werden. Wir nennen (8) und (10) strukturelle und Datenverteilungsform des *EVA Modells in Effektdarstellung mit Referenzgruppe*.

Beweis

Es gilt

$$\mu_i = \mu_0 + \mu_i - \mu_0. \quad (11)$$

Die Parameterisierungen mit μ_i und mit $\mu_0 + \mu_i - \mu_0$ sind also gleich und damit äquivalent. Dann folgt aber auch

$$\mu_i = \mu_0 + (\mu_i - \mu_0) =: \mu_0 + \alpha_i \text{ für } i = 1, \dots, p \quad (12)$$

Mit $\alpha_1 := 0$ gilt dann $\mu_1 = \mu_0$ und $\mu_i = \mu_0 + \alpha_i$ für $i = 2, \dots, p$, wie im Theorem behauptet.

Theorem (EVA Modell in Effektdarstellung mit Referenzgruppe II)

Gegeben sei die strukturelle Form des EVA Modells in Effektdarstellung mit Referenzgruppe. Dann hat dieses Modell die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n), n := \sum_{i=1}^p n_i \quad (13)$$

$$y := \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix} \in \mathbb{R}^n, X := \begin{pmatrix} 1 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 0 \\ 1 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 1 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 1 \end{pmatrix} \in \mathbb{R}^{n \times p}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \text{ und } \sigma^2 > 0.$$

Beweis

Das Theorem ergibt sich direkt mit den Regeln der Matrixmultiplikation.

Modellformulierung

Beispiel

Es seien

$$p = 4 \text{ und } n_i := 3 \text{ für } i = 1, \dots, p, \text{ also } n = 12. \quad (14)$$

Dann gilt

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{12}, \sigma^2 I_{12}) \quad (15)$$

mit

$$y := \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{12 \times 4}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (16)$$

Beispiel

```
# Modellformulierung
library(MASS)
m = 3
p = 4
n = p*m
Xt = cbind(
  matrix(1,nrow = n, ncol = 1),
  kronecker(diag(p), matrix(1,nrow = m,ncol = 1)))
X = Xt[,-2]
I_n = diag(n)
beta = matrix(c(1,2), nrow = p)
sigsqr = 14

# Multivariate Normalverteilung
# Anzahl von Datenpunkten der iten Gruppe
# Anzahl Gruppen
# Gesamtanzahl Datenpunkte
# Designmatrix

# n x n Einheitsmatrix
# \beta = (\mu_0, \alpha_2, \alpha_3, \alpha_4)
# \sigma^2

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n)

# eine Realisierung eines n-dimensionalen ZVs

>      [,1] [,2] [,3] [,4]
> [1,]  1   0   0   0
> [2,]  1   0   0   0
> [3,]  1   0   0   0
> [4,]  1   1   0   0
> [5,]  1   1   0   0
> [6,]  1   1   0   0
> [7,]  1   0   1   0
> [8,]  1   0   1   0
> [9,]  1   0   1   0
> [10,] 1   0   0   1
> [11,] 1   0   0   1
> [12,] 1   0   0   1
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Theorem (Betaparameterschätzung im EVA Modell)

Gegeben sei die Designmatrixform des EVA in Effektdarstellung mit Referenzgruppe Dann ergibt sich für den Betaparameterschätzer

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_p - \bar{y}_1 \end{pmatrix} \quad (17)$$

wobei

$$\bar{y}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad (18)$$

das Stichprobenmittel der i ten Gruppe bezeichnet.

Bemerkungen

- Der Erwartungswertparameter der ersten Gruppe wird durch das Stichprobenmittel der ersten Gruppe geschätzt; der Effektparameter der i ten Gruppe wird durch die Differenz des Stichprobenmittels der i ten und der ersten Gruppe geschätzt. Die Beziehung zwischen dem Varianzparameterschätzer $\hat{\sigma}^2$ und dem Konzept der gepoolten Stichprobenvarianz betrachten wir an dieser Stelle nicht genauer.

Modellschätzung

Beweis

Wir halten zunächst fest, dass

$$X^T X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & & & \vdots & & & & \vdots & \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix}$$
$$= \begin{pmatrix} n & n_2 & n_3 & \dots & n_p \\ n_2 & n_2 & 0 & \dots & 0 \\ n_3 & 0 & n_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_p & 0 & 0 & \dots & n_p \end{pmatrix}.$$

Beweis (fortgeführt)

Die Inverse von $X^T X$ ist

$$(X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & \dots & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \dots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_1} & \frac{1}{n_1} & \dots & \frac{n_1+n_p}{n_1 n_p} \end{pmatrix}. \quad (19)$$

So gilt zum Beispiel für $p = 3$, dass

$$X^T X = \begin{pmatrix} n & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{pmatrix} \text{ und } (X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{1}{n_1} & \frac{n_1+n_3}{n_1 n_3} \end{pmatrix}. \quad (20)$$

Modellschätzung

Beweis (fortgeführt)

Wir halten weiterhin fest, dass

$$X^T y = \begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & & & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix} \quad (21)$$
$$= \begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \\ \sum_{j=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} y_{pj} \end{pmatrix}.$$

Modellschätzung

Beweis (fortgeführt)

Es ergibt sich also

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & \cdots & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \cdots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_1} & \frac{1}{n_1} & \cdots & \frac{n_1+n_p}{n_1 n_p} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \\ \sum_{j=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} y_{pj} \end{pmatrix}. \quad (22)$$

Für die erste Komponente von $\hat{\beta}$ ergibt sich damit

$$\begin{aligned} \hat{\beta}_1 &= \frac{1}{n_1} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} - \frac{1}{n_1} \sum_{j=1}^{n_2} y_{2j} - \cdots - \frac{1}{n_1} \sum_{j=1}^{n_p} y_{pj} \\ &= \frac{1}{n_1} \left(\left(\sum_{j=1}^{n_1} y_{1j} + \sum_{j=1}^{n_2} y_{2j} + \cdots + \sum_{j=1}^{n_p} y_{pj} \right) - \sum_{j=1}^{n_2} y_{2j} - \cdots - \sum_{j=1}^{n_p} y_{pj} \right) \\ &= \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ &= \bar{y}_1. \end{aligned} \quad (23)$$

Modellschätzung

Beweis (fortgeführt)

Für die zweite Komponente von $\hat{\beta}$ und analog für alle weiteren ergibt sich

$$\begin{aligned}\hat{\beta}_2 &= -\frac{1}{n_1} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} + \frac{n_1 + n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} + \frac{1}{n_1} \sum_{j=1}^{n_3} y_{3j} + \cdots + \frac{1}{n_1} \sum_{j=1}^{n_p} y_{pj} \\ &= \frac{n_1 + n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \left(\left(\sum_{j=1}^{n_1} y_{1j} + \sum_{j=1}^{n_2} y_{2j} + \cdots + \sum_{j=1}^{n_p} y_{pj} \right) - \sum_{j=1}^{n_3} y_{3j} - \cdots - \sum_{j=1}^{n_p} y_{pj} \right) \\ &= \frac{n_1 + n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} - \frac{1}{n_1} \sum_{j=1}^{n_2} y_{2j} \\ &= \frac{n_1 + n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ &= \frac{n_1}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ &= \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} = \bar{y}_2 - \bar{y}_1.\end{aligned}$$

Modellschätzung

```
# Dateneinlesen
fname = file.path(getwd(), "10_Daten", "10_Einfaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# Datengruppen
y_1 = D$BDI[D$Condition == "F2F"] # BDI Differenzwerte in der F2F Gruppe
y_2 = D$BDI[D$Condition == "ONL"] # BDI Differenzwerte in der ONL Gruppe
y_3 = D$BDI[D$Condition == "WLC"] # BDI Differenzwerte in der ONL Gruppe

# Modellformulierung
p = 3 # drei Gruppen
m = length(y_1) # balancierteres Design mit n_i = 40
n = p*m # Datenvektordimension
y = matrix(c(y_1, y_2, y_3), nrow = n) # Datenvektor
Xt = cbind( # Designmatrix
  matrix(1, nrow = n, ncol = 1),
  kronecker(diag(p), matrix(1, nrow = m, ncol = 1)))
X = Xt[,-2]

# Modellschätzung
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = t(eps_hat) %*% eps_hat / (n-p) # Varianzparameterschätzer
s_sqr_123 = ((m-1)*var(y_1) + # gepoolte Stichprobenvarianz
  (m-1)*var(y_2) +
  (m-1)*var(y_3))/(m+m+m-p)

# Ausgabe
cat("hat{beta} : ", beta_hat,
  "\nbar{y}_1,bar{y}_2,bar{y}_3 : ", c(mean(y_1),mean(y_2),mean(y_3)),
  "\nbar{y}_1,bar{y}_2-bar{y}_1,bar{y}_3-bar{y}_1 : ", c(mean(y_1),mean(y_2)-mean(y_1),mean(y_3)-mean(y_1)),
  "\nhat{sigsqr} : ", sigsqr_hat,
  "\ns_123^2 : ", s_sqr_123,
  "\ns_y^2 : ", var(y))

> hat{beta} : 5.28 0.65 -5.8
> bar{y}_1,bar{y}_2,bar{y}_3 : 5.28 5.92 -0.525
> bar{y}_1,bar{y}_2-bar{y}_1,bar{y}_3-bar{y}_1 : 5.28 0.65 -5.8
> hat{sigsqr} : 18.4
> s_123^2 : 18.4
> s_y^2 : 26.6
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Vorbemerkungen und Überblick

Prinzipiell sind alle Parameterschätzwerte in einem EVA Modell von Interesse; mit der T-Teststatistik können einzelne oder lineare Kombinationen der Betaparameterschätzwerte im Sinne von Hypothesentests evaluiert werden.

Nichtsdestotrotz steht im EVA Szenario häufig die Evaluation der Nullhypothese, dass alle Effektparameter gleich Null sind im Vordergrund. Intuitiv besagt diese Nullhypothese, dass die wahren, aber unbekannt, Erwartungswertparameter aller Gruppen identisch sind, formal haben wir

$$H_0 : \alpha_i = 0 \text{ für alle } i = 2, \dots, p \Leftrightarrow \Theta_0 := \mathbb{R} \times \{0_{p-1}\} \quad (25)$$

und

$$H_1 : \alpha_i \neq 0 \text{ für mindestens ein } i = 2, \dots, p \Leftrightarrow \Theta_1 := \mathbb{R}^p \setminus \mathbb{R} \times \{0_{p-1}\} \quad (26)$$

Zur Überprüfung von H_0 wird im Allgemeinen eine F -Teststatistik eingesetzt.

Im Folgenden entwickeln wir zunächst diese F -Teststatistik anhand einer *Quadratsummenzerlegung* der Datenvariabilität in einem EVA Szenario und betrachten in diesem Zusammenhang auch das dem R^2 analoge Maß η^2 der Effektstärke im EVA Szenario. Ausgestattet mit der speziellen Form der F -Teststatistik in dem hier betrachteten Szenario diskutieren wir dann den traditionellen EVA Hypothesentest. Wir verzichten auf eine Analyse der Testgütfunktion und eine Diskussion der Powerfunktion.

Theorem (Quadratsummenzerlegung bei einfaktorieller Varianzanalyse)

Für $i = 1, \dots, p$ und $j = 1, \dots, n_i$ sei y_{ij} die j te Datenvariable in der i ten Gruppe eines EVA Szenarios. Weiterhin seien mit $n := \sum_{i=1}^p n_i$

$$\bar{y} := \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \quad \text{das Gesamtstichprobenmittel}$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad \text{das } i\text{te Stichprobenmittel}$$

sowie

$$\text{SQT} := \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \quad \text{die Total Sum of Squares}$$

$$\text{SQB} := \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2 \quad \text{die Between Sum of Squares}$$

$$\text{SQW} := \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad \text{die Within Sum of Squares}$$

Dann gilt

$$\text{SQT} = \text{SQB} + \text{SQW}. \quad (27)$$

Bemerkung

- In Analogie zur Quadratsummenzerlegung bei einer Ausgleichsgerade (vgl. Einheit (3) Korrelation) wird die *Within Sum of Squares* auch als *Residual Sum of Squares* bezeichnet.

Beweis

Es gilt

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i + \bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} \left((y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}) \right)^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} \left((y_{ij} - \bar{y}_i)^2 + 2(y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}) + (\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{j=1}^{n_i} 2(y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}) + \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 \right) \end{aligned} \tag{28}$$

Beweis (fortgeführt)

und weiter

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) + n_i(\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \sum_{j=1}^{n_i} \left(y_{ij} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \left(\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \right) \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \left(\sum_{j=1}^{n_i} y_{ij} - \frac{n_i}{n_i} \sum_{j=1}^{n_i} y_{ij} \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \end{aligned}$$

Beweis (fortgeführt)

und weiter

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \left(\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} y_{ij} \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + n_i(\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^p n_i(\bar{y}_i - \bar{y})^2 \\ &= \text{SQW} + \text{SQB} \end{aligned} \tag{29}$$

und damit direkt

$$\text{SQT} = \text{SQB} + \text{SQW}. \tag{30}$$

Definition (Effektstärkenmaß η^2)

Für ein EVA Szenario seien die Between Sum of Squares SQB und die Total Sum of Squares SQT definiert wie oben. Dann ist das *Effektstärkenmaß* η^2 definiert als

$$\eta^2 := \frac{\text{SQB}}{\text{SQT}} \quad (31)$$

Bemerkungen

- η^2 ist analog zum Bestimmtheitsmaß R^2 der Regression definiert.
- η^2 gibt den Anteil der Varianz zwischen den Gruppen an der Gesamtvarianz der Daten an.
- Mit dem Theorem zur Quadratsummenzerlegung bei EVA folgt sofort $0 \leq \eta^2 \leq 1$, da

$$\begin{aligned} \text{SQB} = 0 &\Rightarrow \text{SQT} = \text{SQW} \text{ und } \eta^2 = 0 \\ \text{SQW} = 0 &\Rightarrow \text{SQT} = \text{SQB} \text{ und } \eta^2 = 1 \end{aligned} \quad (32)$$

Theorem (F-Teststatistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (33)$$

die Designmatrixform der Effektdarstellung mit Referenzgruppe des EVA Modells und im Sinne der Definition der F-Statistik (vgl. Einheit (7) Modellevaluation) sei dieses Modell partitioniert mit $p_1 := 1$ und $p_2 := p - 1$. Weiterhin seien

$$\text{MSB} := \frac{\text{SQB}}{p-1} \quad \text{die Mean Between Sum of Squares}$$

$$\text{MSW} := \frac{\text{SQW}}{n-p} \quad \text{die Mean Within Sum of Squares}$$

respektive. Dann gilt

$$F = \frac{\text{MSB}}{\text{MSW}}. \quad (34)$$

Bemerkungen

- $p_1 := 1$ impliziert, dass das reduzierte Modell die Designmatrix $X_1 := 1_n$ hat.
- $p_1 := 1$ impliziert zudem, dass das reduzierte Modell den Betaparameter $\beta := \mu_0$ hat.
- $p_1 := 1$ impliziert damit auch, dass das reduzierte Modell keine Effektparameter hat.
- Die Zahl $p - 1$ wird auch als "Between Freiheitsgrade" bezeichnet.
- Die Zahl $n - p$ wird auch als "Within Freiheitsgrade" bezeichnet.

Beweis

Wir halten zunächst fest, dass für den Betaparameterschätzer des reduzierten Modells gilt, dass

$$\hat{\beta}_1 = (X_1^T X_1)^{-1} X_1^T y = (1_n^T 1_n)^{-1} 1_n^T y = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} = \bar{y}. \quad (35)$$

Weiterhin ergibt sich

$$\hat{\varepsilon}_1^T \hat{\varepsilon}_1 = (y - X_1 \hat{\beta}_1)^T (y - X_1 \hat{\beta}_1) = (y - 1_n \bar{y})^T (y - 1_n \bar{y}) = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \text{SQT}. \quad (36)$$

Der Betaparameterschätzer des vollständigen Modells ergibt sich wie oben gesehen zu

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_m \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \vdots \\ \frac{1}{n_m} \sum_{j=1}^{n_m} y_{mj} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_p - \bar{y}_1 \end{pmatrix}, \quad (37)$$

so dass

Beweis (fortgeführt)

$$\begin{aligned} \hat{\varepsilon}^T \hat{\varepsilon} &= (y - X\hat{\beta})^T (y - X\hat{\beta}) \\ &= \left(\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix} - \begin{pmatrix} 1 & 0 & & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 0 \\ 1 & 1 & & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 1 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 1 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_p - \bar{y}_1 \end{pmatrix} \right)^T (y - X\hat{\beta}) \end{aligned}$$

Modellevaluation

Beweis (fortgeführt)

und weiter

$$\begin{aligned} \hat{\epsilon}^T \hat{\epsilon} &= \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{2n_2} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{p1} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \\ \vdots \\ y_{pn_p} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{2n_2} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{p1} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \\ \vdots \\ y_{pn_p} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \end{pmatrix} = \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \\ \vdots \\ y_{p1} - \bar{y}_p \\ \vdots \\ y_{pn_p} - \bar{y}_p \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \\ \vdots \\ y_{p1} - \bar{y}_p \\ \vdots \\ y_{pn_p} - \bar{y}_p \end{pmatrix} \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \\ &= \text{SQW}. \end{aligned}$$

Beweis (fortgeführt)

Mit dem Theorem zur Quadratsummenzerlegung bei einfaktorieller Varianzanalyse

$$SQT = SQB + SQW \Leftrightarrow SQB = SQT - SQW \quad (38)$$

folgt sofort, dass

$$\begin{aligned} SQB &= SQT - SQW \\ &= \hat{\epsilon}_1^T \hat{\epsilon}_1 - \hat{\epsilon}^T \hat{\epsilon}. \end{aligned} \quad (39)$$

Dann aber folgt auch direkt, dass

$$\begin{aligned} \frac{MSB}{MSW} &= \frac{\frac{SQB}{p-1}}{\frac{SQW}{n-p}} \\ &= \frac{\frac{\hat{\epsilon}_1^T \hat{\epsilon}_1 - \hat{\epsilon}^T \hat{\epsilon}}{p-1}}{\frac{\hat{\epsilon}^T \hat{\epsilon}}{n-p}} \\ &= F \end{aligned} \quad (40)$$

Theorem (Effektstärkenmaß η^2 und F-Teststatistik)

Für ein EVA Szenario mit p Gruppen und Gesamtdatenpunktzahl n seien das Effektstärkenmaß η^2 und die F -Teststatistik wie oben definiert. Dann gilt

$$\eta^2 = \frac{F(p-1)}{F(p-1) + (n-p)} \quad (41)$$

Bemerkungen

- Das Verhältnis von F und η^2 ist Analog zum Verhältnis von T und Cohen's d .
- Die gleichzeitige Angabe von F und η^2 ist redundant.

Modellevaluation

Beweis

Wir halten zunächst fest, dass

$$F = \frac{SQB}{SQW} \cdot \frac{n-p}{p-1} \Leftrightarrow SQB = \frac{p-1}{n-p} \cdot SQW \cdot F \quad (42)$$

Damit folgt dann

$$\begin{aligned} \eta^2 &= \frac{SQB}{SQT} = \frac{SQB}{SQB + SQW} = \frac{\frac{p-1}{n-p} \cdot SQW \cdot F}{\frac{p-1}{n-p} \cdot SQW \cdot F + SQW} = \frac{\frac{F(p-1)}{n-p} \cdot SQW}{\frac{F(p-1)}{n-p} \cdot SQW + SQW} \\ &= \frac{\frac{F(p-1)}{n-p} \cdot SQW}{\left(\frac{F(p-1)}{n-p} + 1\right) \cdot SQW} \\ &= \frac{\frac{F(p-1)}{n-p}}{\frac{F(p-1)}{n-p} + \frac{n-p}{n-p}} \\ &= \frac{\frac{F(p-1)}{n-p}}{\frac{F(p-1) + (n-p)}{n-p}} \\ &= \frac{F(p-1)}{F(p-1) + (n-p)} \end{aligned} \quad (43)$$

Gliederung (vgl. WTFI Einheiten (12) - (14))

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik ✓
- (4) Test
- (5) Testumfangkontrolle
- (6) p-Werte

(4) Test

Definition (Einfaktorielle Varianzanalyse F-Test)

Gegeben sei das Modell der einfaktoriellen Varianzanalyse sowie die zusammengesetzten Null- und Alternativhypothesen

$$H_0 : \alpha_i = 0 \text{ für alle } i = 2, \dots, p \Leftrightarrow \Theta_0 := \mathbb{R} \times \{0_{p-1}\} \quad (44)$$

und

$$H_1 : \alpha_i \neq 0 \text{ für mindestens ein } i = 2, \dots, p \Leftrightarrow \Theta_1 := \mathbb{R}^p \setminus \mathbb{R} \times \{0_{p-1}\} \quad (45)$$

respektive. Weiterhin sei die F-Teststatistik definiert durch

$$F = \frac{MSB}{MSW} \quad (46)$$

mit der Mean Sum of Squares Between MSB und der Mean Sum of Square Within MSW definiert wie oben. Dann ist der *einfaktoriellen Varianzanalyse F-Test (EVA F-Test)* definiert als der kritische Wert-basierte Test

$$\phi(y) := 1_{\{F \geq k\}} := \begin{cases} 1 & F \geq k \\ 0 & F < k \end{cases} . \quad (47)$$

(5) Testumfangkontrolle

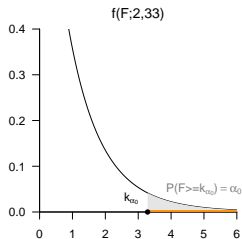
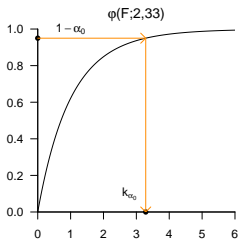
Theorem (Testumfangkontrolle)

ϕ sei der F-Test zur einfaktoriiellen Varianzanalyse. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \varphi^{-1}(1 - \alpha_0; p - 1, n - p), \quad (48)$$

wobei $\varphi^{-1}(\cdot; p - 1, n - p)$ die inverse KVF der f -Verteilung mit Freiheitsgradparametern $p - 1$ und $n - p$ ist.

Wahl von $k_{\alpha_0} := \varphi^{-1}(1 - \alpha_0; p - 1, n - p)$ mit $p = 3, n = 12, \alpha_0 := 0.05$ und Ablehnungsbereich



Beweis

Die Testgütefunktion des betrachteten Tests im vorliegenden Testscenario ist definiert als

$$q : \mathbb{R} \rightarrow [0, 1], \beta \mapsto q_\phi(\beta) := \mathbb{P}_\beta(\phi = 1). \quad (49)$$

Wir haben in Einheit (7) Modellevaluation gesehen, dass die F-Teststatistik für $p_2 = p - 1$ nach einer nichtzentralen f -Verteilung verteilt ist,

$$F \sim f(\delta, p - 1, n - p). \quad (50)$$

Weiterhin ist der Ablehnungsbereich des hier betrachteten Tests gegeben als $[k, \infty[$. Für die funktionale Form der Testgütefunktion ergibt sich also

$$\begin{aligned} \mathbb{P}_\beta(\phi = 1) &= \mathbb{P}_\beta(F \in [k, \infty[) \\ &= \mathbb{P}_\beta(F \geq k) \\ &= 1 - \mathbb{P}_\beta(F \leq k) \\ &= 1 - \varphi(k; \delta, p - 1, n - p), \end{aligned} \quad (51)$$

wobei $\varphi(k; \delta, p - 1, n - p)$ den Wert der KVF der nichtzentralen f -Verteilung an der Stelle k und mit Nichtzentralitätsparameter δ sowie Freiheitsgradparametern $p - 1$ und $n - p$ bezeichnet (vgl. Einheit (7) Modellevaluation).

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich gelten, dass

$$q_\phi(\beta) \leq \alpha_0 \text{ für alle } \beta \in \Theta_0 \text{ mit } \Theta_0 = \mathbb{R} \times \{0_{p-1}\}. \quad (52)$$

Modellevaluation

Beweis

Mit der Form des Nichtzentralitätsparameters (vgl. Einheit (7) Modellevaluation) gegeben durch

$$\delta = \frac{1}{\sigma^2} K^T \beta \left(K^T (X^T X)^{-1} K \right)^{-1} K^T \beta \quad (53)$$

folgt mit $\beta \in \Theta_0$ aus

$$K = \begin{pmatrix} 0 \\ 1_{p-1} \end{pmatrix} \in \mathbb{R}^p \text{ und } \beta = \begin{pmatrix} \mu_0 \\ 0_{p-1} \end{pmatrix} \in \mathbb{R}^p \quad (54)$$

dann aber $\delta = 0$ und somit

$$q_\phi(\beta) = 1 - \varphi(k; p-1, n-p) \text{ für alle } \beta \in \Theta_0. \quad (55)$$

wobei $\varphi(k; p-1, n-p)$ den Wert der KVF der f -Verteilung an der Stelle k mit Freiheitsgradparametern $p-1$ und $n-p$ bezeichnet (vgl. Einheit (7) Modellevaluation). Der Testumfang des betrachteten Tests ergibt sich nach Definition (vgl. WTFI Einheit (12) Hypothesentests) als

$$\alpha = \max_{\beta \in \Theta_0} q_\phi(\beta) = 1 - \varphi(k; p-1, n-p), \quad (56)$$

da $q_\phi(\beta)$ für $\beta \in \Theta_0$ nicht von μ_0 abhängt. Wir müssen also lediglich zeigen, dass die Wahl von k_{α_0} wie im Theorem garantiert, dass ϕ den Testumfang α_0 hat. Sei also $k := k_{\alpha_0}$. Dann gilt für alle $\beta \in \Theta_0$

$$q_\phi(\beta) = 1 - \varphi(\varphi^{-1}(1 - \alpha_0; p-1, n-p); p-1, n-p) = 1 - 1 - \alpha_0 = \alpha_0 \quad (57)$$

und damit ist alles gezeigt.

(7) p-Wert

Nach Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegenden Wert der Teststatistik ablehnen würde. Wir wollen einen vorliegenden Wert der F -Teststatistik hier mit f bezeichnen.

Bei $F = f$ würde H_0 für jedes α_0 mit $f \geq \psi^{-1}(1 - \alpha_0; p - 1, n - p)$ abgelehnt werden. Für ein solches α_0 gilt aber

$$\alpha_0 \geq \mathbb{P}(F \geq f), \quad (58)$$

denn

$$\begin{aligned} f &\geq \psi^{-1}(1 - \alpha_0; p - 1, n - p) \\ \Leftrightarrow \psi(f; p - 1, n - p) &\geq \psi(\psi^{-1}(1 - \alpha_0; p - 1, n - p), p - 1, n - p) \\ \Leftrightarrow \psi(f; p - 1, n - p) &\geq 1 - \alpha_0 \\ \Leftrightarrow \mathbb{P}(F \leq f) &\geq 1 - \alpha_0 \\ \Leftrightarrow \alpha_0 &\geq 1 - \mathbb{P}(F \leq f) \\ \Leftrightarrow \alpha_0 &\geq \mathbb{P}(F \geq f) \end{aligned} \quad (59)$$

Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq \mathbb{P}(F \geq f)$ ist dann $\alpha_0 = \mathbb{P}(F \geq f)$, also folgt

$$\text{p-Wert} = \mathbb{P}(F \geq f) = 1 - \varphi(f; p - 1, n - p). \quad (60)$$

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz von p Gruppendatensätzen $v_{11}, \dots, v_{1n_1}, v_{21}, \dots, v_{2n_2}, \dots, v_{p1}, \dots, v_{pn_p}$ Realisationen von $y_{1j} \sim N(\mu_0, \sigma^2)$ und $y_{ij} \sim N(\mu_0 + \alpha_i, \sigma^2)$ für $i = 2, \dots, p$ mit unbekanntem Parametern $\mu_0, \alpha_i, i = 2, \dots, p$ und $\sigma^2 > 0$ sind.
- Man möchte entscheiden ob $H_0 : \alpha_i = 0$ für alle $i = 2, \dots, p$ eher zutrifft oder eher nicht.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05, p = 3, m = 12, i = 1, 2, 3$ und somit $n = 36$, dass $k_{\alpha_0} = \varphi^{-1}(1 - 0.05; 2, 33) \approx 3.28$ ist.
- Anhand der MSB und MSW berechnet man den realisierten Wert der F-Teststatistik, den wir hier mit f bezeichnen.
- Wenn f größer gleich k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht.
- Die oben entwickelte Theorie garantiert dann, dass man im Mittel in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.
- Schließlich ergibt sich der assoziierte p-Wert der realisierten F-Teststatistik \tilde{F} zu

$$\text{p-Wert} = \mathbb{P}(F \geq f) = 1 - \varphi(f; p - 1, n - p) \quad (61)$$

Modellevaluation

Anwendungsbeispiel

```
# Dateneinlesen
fname = file.path(getwd(), "10_Daten", "10_Einfaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# Datengruppen
y_1 = D$BDI[D$Condition == "F2F"] # BDI Differenzwerte in der F2F Gruppe
y_2 = D$BDI[D$Condition == "ONL"] # BDI Differenzwerte in der ONL Gruppe
y_3 = D$BDI[D$Condition == "WLC"] # BDI Differenzwerte in der ONL Gruppe

# Modellformulierung
p = 3 # drei Gruppen
m = length(y_1) # balancierteres Design mit n_i = 40
n = p*m # Datenvektordimension
y = matrix(c(y_1, y_2, y_3), nrow = n) # Datenvektor
Xt = cbind( # Designmatrix vollständiges Modell
  matrix(1, nrow = n, ncol = 1),
  kronecker(diag(p),
    matrix(1, nrow = m, ncol = 1)))
X = Xt[,-2]
X_1 = X[,1] # Designmatrix reduziertes Modell

# F-Teststatistikevaluation
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer vollständiges Modell
beta_hat_1 = solve(t(X_1) %*% X_1) %*% t(X_1) %*% y # Betaparameterschätzer reduziertes Modell
eps_hat = y - X %*% beta_hat # Residuenvektor vollständiges Modell
eps_hat_1 = y - X_1 %*% beta_hat_1 # Residuenvektor reduziertes Modell
SQT = t(eps_hat_1) %*% eps_hat_1 # Sum of Squares Total
SQW = t(eps_hat) %*% eps_hat # Sum of Squares Within
SQB = SQT - SQW # Sum of Squares Between
DFB = p - 1 # Between Degrees of Freedom
DFW = n - p # Within Degrees of Freedom
DFB = p - 1 # Between Degrees of Freedom
MSB = SQB/DFB # Mean Sum of Squares Between
MSW = SQW/DFW # Mean Sum of Squares Within
Eff = MSB/MSW # F-Teststatistik
p = 1 - pf(Eff, p-1, n-p) # p-Wert
```


Anwendungsbeispiel

```
# Ausgabe  
cat( "DFB :", DFB,  
     "\nDFW :", DFW,  
     "\nSQB :", SQB,  
     "\nSQW :", SQW,  
     "\nMSB :", MSB,  
     "\nMSW :", MSW,  
     "\nF   :", Eff,  
     "\np   :", paste(p))
```

```
> DFB : 2  
> DFW : 117  
> SQB : 1009  
> SQW : 2153  
> MSB : 504  
> MSW : 18.4  
> F   : 27.4  
> p   : 1.7192636203589e-10
```

Anwendungsbeispiel

```
# Dateneinlesen
fname = file.path(getwd(), "10_Daten", "10_Einfaktorielle_Varianzanalyse_Daten.csv")
D      = read.table(fname, sep = ",", header = TRUE)

# Benutzung von R's aov Funktion und Ausgabe
res.aov = aov(D$BDI ~ D$Condition, data = D)
summary(res.aov)

>
      Df Sum Sq Mean Sq F value Pr(>F)
> D$Condition  2   1009     504   27.4 1.7e-10 ***
> Residuals  117   2153      18
> ---
> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario einer einfaktoriellen Varianzanalyse (EVA).
2. Geben Sie die Definition des EVA Modells in Erwartungswertparameterdarstellung wieder.
3. Geben Sie die strukturelle Form des EVA Modells in Effektdarstellung wieder.
4. Erläutern Sie die Motivation für die Reparameterisierung des EVA Modells
5. Welche Bedeutung haben die Parameter $\mu_0, \alpha_2, \dots, \alpha_p$ in der Effektparameterdarstellung des EVA Modells?
6. Warum gibt es bei p Gruppen eines EVA Szenarios nur die $p - 1$ Effektparameter $\alpha_2, \dots, \alpha_p$?
7. Geben Sie die Designmatrixform des EVA Modells in Effektdarstellung wieder.
8. Formulieren Sie die Designmatrix eines EVA Modells mit $n_i = 3$ und $p = 2$.
9. Formulieren Sie die Designmatrix eines EVA Modells mit $n_i = 2$ und $p = 5$.
10. Geben Sie das Theorem zur Betaparameterschätzung im EVA Modell wieder.
11. Mit welchen deskriptiven Statistiken werden die Parameter $\mu_0, \alpha_2, \dots, \alpha_p$ eines EVA Modells geschätzt?
12. Geben Sie das Theorem zur Quadratsummenzerlegung bei einfaktorieller Varianzanalyse wieder.
13. Erläutern Sie die Begriffe Total Sum of Squares, Between Sum of Squares, Within Sum of Squares der EVA.
14. Geben Sie die Definition des Effektstärkenmaßes η^2 an.

Selbstkontrollfragen

15. Wann nimmt das Effektstärkenmaß η^2 der EVA seinen Minimalwert an und wie lautet dieser?
16. Wann nimmt das Effektstärkenmaß η^2 der EVA seinen Maximalwert an und wie lautet dieser?
17. Geben Sie das Theorem zur F-Teststatistik der EVA wieder.
18. Erläutern Sie die Begriffe Mean Between Sum of Squares und Mean Within Sum of Squares der EVA.
19. Geben Sie das Theorem zum Zusammenhang von Effektstärkenmaß η^2 und F-Teststatistik der EVA wieder.
20. Geben Sie die Definition des EVA F-Test wieder.
21. Erläutern sie die Null- und Alternativhypothesen des EVA F-Tests.
22. Geben Sie das Theorem zur Testumfangkontrolle der EVA wieder.
23. Skizzieren Sie den Beweis zur Testumfangkontrolle der EVA.
24. Geben Sie den p-Wert zum F-Test der EVA wieder.
25. Betrachten Sie die Daten zur Therapiedauer der Patient:innen in der Face-to-Face, Online, und Waitlist Control Gruppe im Beispieldatensatz. Erstellen Sie ein Balkendiagramme der Gruppenmittelwerte und Gruppenstandardabweichungen und erläutern Sie dieses. Führen Sie zu diesen Daten einen EVA F-Test mit $\alpha := 0.05$ durch. Dokumentieren Sie ihre Ergebnisse. Was folgern Sie aus den sich ergebenden Resultaten?

Georgii, Hans-Otto. 2009. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 4., überarb. und erw. Aufl. De-Gruyter-Lehrbuch. Berlin: de Gruyter.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

Datum	Einheit	Thema
08.04.2022	Grundlagen	(1) Regression
	Osterpause	
22.04.2022	Grundlagen	(2) Korrelation
29.04.2022	Grundlagen	(3) Matrizen
06.05.2022	Grundlagen	(4) Normalverteilungen
13.05.2022	Theorie	(5) Modellformulierung
20.05.2022	Theorie	(6) Modellschätzung
27.05.2022	Theorie	(7) Modellevaluation
03.06.2021	Anwendung	(8) Studiendesign
10.06.2021	Anwendung	(9) T-Tests
17.06.2021	Anwendung	(10) Einfaktorielle Varianzanalyse
24.06.2022	Anwendung	(11) Zweifaktorielle Varianzanalyse
01.07.2022	Anwendung	(12) Multiple Regression (9 - 12 Uhr)
01.07.2022	Anwendung	(13) Kovarianzanalyse (13 - 15 Uhr)
14.07.2022	Klausurtermin	
März 2023	Klausurwiederholungstermin	

(11) Zweifaktorielle Varianzanalyse

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Randomisiertes zweifaktorielles Studiendesigns mit crossed design

⇒ Zweifaktorielle Varianzanalyse (ZVA) = Two-way Analysis of Variance (ANOVA)

- Eine univariate abhängige Variable bestimmt an randomisierten experimentellen Einheiten.
- Zwei diskrete unabhängige Variablen, die mindestens zweistufig sind.
- Die unabhängigen Variablen werden **Faktoren** genannt.
- Die Stufen der Faktoren werden auch **Faktorlevel** genannt.
- Jedes Level eines Faktors wird mit allen Level des anderen Faktors kombiniert.
- Die Kombination zweier spezifischer Faktorlevel wird **Zelle** des Designs genannt.

Zweifaktorielle Studiendesigns werden üblicherweise anhand ihrer Faktorlevel bezeichnet

2 × 2 ANOVA : Faktor A mit Level 1,2	Faktor B mit Level 1,2
2 × 3 ANOVA : Faktor A mit Level 1,2	Faktor B mit Level 1,2,3
4 × 2 ANOVA : Faktor A mit Level 1,2,3,4	Faktor B mit Level 1,2
3 × 1 ANOVA : Faktor A mit Level 1,2,3	Faktor B mit Level 1

Generell sind 2 × 2 Designs sehr populär, wir fokussieren auf diesen Fall.

Die Zellen eines 2 × 2 Designs werden im Folgenden auch als Gruppen bezeichnet.

Konzeptuelles Design

		Faktor B	
		Level 1	Level 2
Faktor A	Level 1	A1B1	A1B2
	Level 2	A2B1	A2B2

Die Zellen/Gruppen des Designs sind hier mit A1B1, A1B2, A2B1 und A2B2 bezeichnet.

Datennotation

		Faktor B	
		Level 1	Level 2
Faktor A	Level 1	$y_{111} = 1.2$ $y_{112} = 0.4$ $y_{113} = 1.7$ \vdots $y_{11n_{11}} = 2.1$	$y_{121} = 4.1$ $y_{122} = 1.8$ $y_{123} = 3.3$ \vdots $y_{12n_{12}} = 5.9$
	Level 2	$y_{211} = 0.1$ $y_{212} = 0.8$ $y_{213} = 2.7$ \vdots $y_{21n_{21}} = 1.4$	$y_{221} = 7.4$ $y_{222} = 6.2$ $y_{223} = 9.5$ \vdots $y_{22n_{22}} = 6.1$

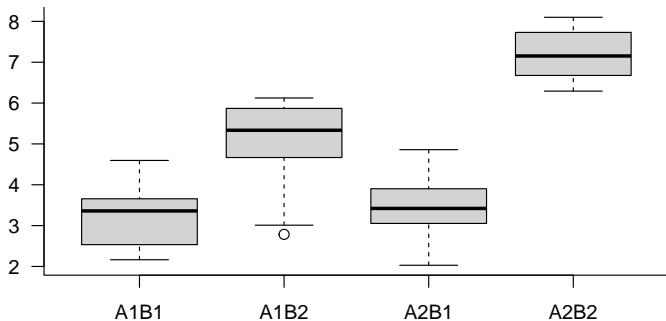
y_{ijk} bezeichnet die Datenvariable der k ten experimentellen Einheit ($k = 1, \dots, n_{ij}$) im i ten Level von Faktor A und j ten Level von Faktor B ($i = 1, 2, j = 1, 2$)

Datentabellenform mit $n_{ij} := 12, 1 \leq i, j \leq 2$

	y_{11k}	y_{12k}	y_{21k}	y_{22k}
1	2.37	4.38	4.12	6.61
2	3.18	2.79	3.44	6.94
3	2.16	6.12	3.34	8.10
4	4.59	4.96	2.03	7.76
5	3.33	4.98	3.02	6.83
6	2.18	5.94	3.92	6.75
7	3.49	5.82	4.86	7.70
8	3.74	5.59	3.40	7.56
9	3.58	5.92	3.89	6.31
10	2.69	5.78	3.45	6.29
11	4.51	5.08	2.12	7.37
12	3.39	3.01	3.08	7.77

y_{ijk} bezeichnet die Datenvariable der k ten experimentellen Einheit ($k = 1, \dots, n_{ij}$) im i ten Level von Faktor A und j ten Level von Faktor B ($i = 1, 2, j = 1, 2$)

Datendeskription durch Boxplot



Haupteffekte (main effects) und Interaktionen (interactions)

Man unterscheidet intuitiv hinsichtlich der Gruppenmittelwerte *Haupteffekte* und *Interaktionen*

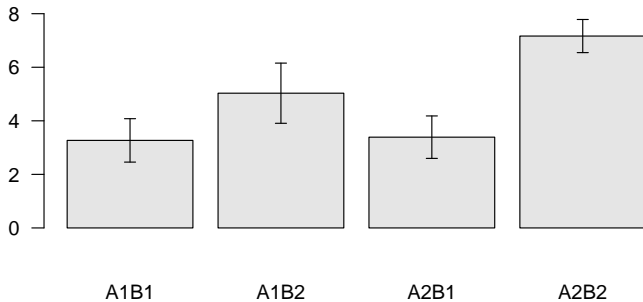
- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor A*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor A, jeweils gemittelt über die zwei Level von Faktor B, unterscheiden.
- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor B*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor B, jeweils gemittelt über die zwei Level von Faktor A, unterscheiden.
- Intuitiv spricht man vom Vorliegen einer *Interaktion der Faktoren A und B*, wenn der Unterschied der Gruppenmittelwerte von Faktor A zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor B ausgeprägt ist bzw. wenn der Unterschied der Gruppenmittelwerte von Faktor B zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor A ausgeprägt ist.

Intuitiv beziehen sich Haupteffekte also auf (marginale) Unterschiede (Differenzen), während sich Interaktionen auf Unterschiede von Unterschieden (Differenzen von Differenzen) beziehen.

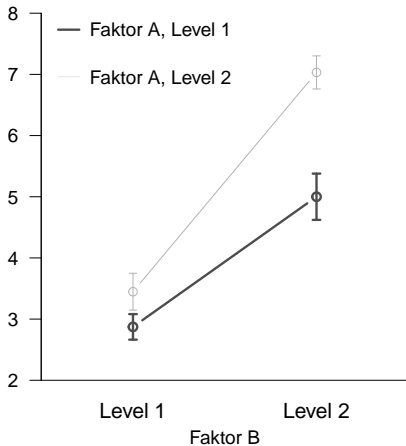
Das Vorhandensein einer Interaktion besagt lediglich, dass sich die Unterschiede der Gruppenmittelwerte zwischen den Leveln eines experimentellen Faktors in Abhängigkeit von den Leveln des anderen experimentellen Faktors ändern, es macht aber keine Aussage darüber, warum dies so ist.

⇒ Haupteffekte und Interaktionen sind Datenmuster, keine wissenschaftlichen Theorien.

Datendeskription durch Barplot



Datendeskription durch Lineplot



Anwendungsbeispiel

BDI Datenanalyse bei zwei Therapiearten für zwei Altersklassen

- Faktor (A) Therapie mit Level (1) Face-to-face (F2F) und Level (2) Online (ONL)
- Faktor (B) Altersklasse mit Level (1) Young Adults (YA) und (2) und Older Adults (OA)
- $n_{ij} := 20, 1 \leq i, j \leq 2$ Patient:innen in jeder Bedingung, $n = 80$ Patient:innen insgesamt.

2×2 randomisiertes zweifaktorielles Studiendesigns $\Rightarrow 2 \times 2$ ANOVA

- Inferentielle Evidenz für einen Haupteffekt der Therapieart?
- Inferentielle Evidenz für einen Haupteffekt der Altersklasse?
- Inferentielle Evidenz für Interaktion (Verschiedene Therapieeffekte zwischen Altersklassen)?

Anwendungsszenario

Anwendungsbeispiel

```
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)
```

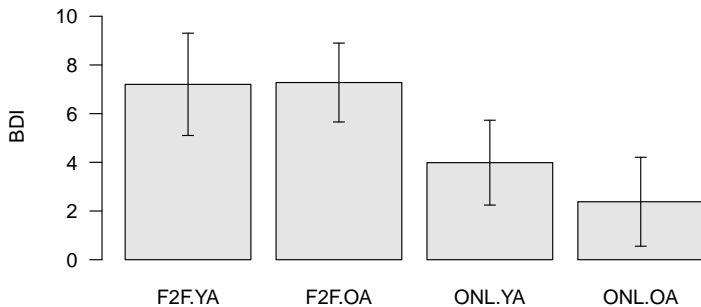
	ID	Therapy	Age	BDI
1	1	F2F	YA	5.82
2	2	F2F	YA	7.15
3	3	F2F	YA	7.00
4	4	F2F	YA	6.11
5	5	F2F	YA	7.58
6	6	F2F	YA	4.49
21	21	F2F	OA	6.73
22	22	F2F	OA	8.14
23	23	F2F	OA	4.91
24	24	F2F	OA	6.27
25	25	F2F	OA	10.96
26	26	F2F	OA	9.87
41	41	ONL	YA	5.53
42	42	ONL	YA	6.20
43	43	ONL	YA	3.88
44	44	ONL	YA	3.21
45	45	ONL	YA	3.17
46	46	ONL	YA	1.25
61	61	ONL	OA	3.19
62	62	ONL	OA	3.64
63	63	ONL	OA	3.89
64	64	ONL	OA	1.97
65	65	ONL	OA	1.91
66	66	ONL	OA	4.25

Anwendungsbeispiel

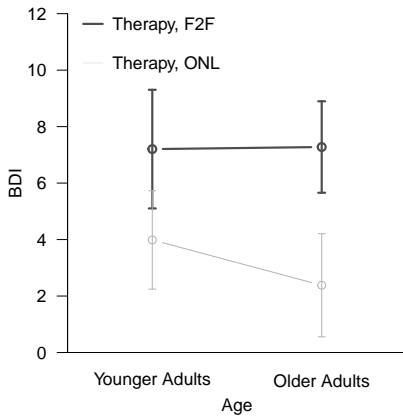
```
# Datenreformatierung
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D      = read.table(fname, sep = ",", header = TRUE)
A1B1  = D$BDI[D$Therapy == "F2F" & D$Age == "YA" ]
A1B2  = D$BDI[D$Therapy == "F2F" & D$Age == "OA" ]
A2B1  = D$BDI[D$Therapy == "ONL" & D$Age == "YA" ]
A2B2  = D$BDI[D$Therapy == "ONL" & D$Age == "OA" ]
Y      = data.frame("F2F YA" = A1B1, "F2F OA" = A1B2, "ONL YA" = A2B1, "ONL OA" = A2B2)
```

	F2F.YA	F2F.OA	ONL.YA	ONL.OA
1	5.82	6.73	5.526	3.188
2	7.15	8.14	6.200	3.642
3	7.00	4.91	3.881	3.888
4	6.11	6.27	3.211	1.968
5	7.58	10.96	3.170	1.910
6	4.49	9.87	1.246	4.250
7	5.13	4.74	3.892	-2.429
8	8.22	7.68	4.775	0.758
9	5.58	5.78	3.794	2.780
10	7.95	7.80	6.717	5.024
11	11.35	8.76	4.836	1.389
12	7.31	6.78	3.044	3.152
13	9.93	8.54	1.058	3.477
14	3.39	7.73	3.688	2.975
15	7.38	5.58	3.888	0.359
16	5.51	5.62	5.240	2.659
17	7.06	8.11	0.021	5.191
18	8.38	8.39	4.149	0.329
19	6.92	6.49	5.564	2.367
20	11.80	6.67	5.838	0.747

Datendeskription durch Barplot



Datendeskription durch Lineplot



Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Modell der additiven ZVA

In Analogie zur einfaktoriellen Varianzanalyse (EVA) möchte man in der additiven ZVA die Gruppenerwartungswerte μ_{ij} mit $i = 1, \dots, I$ für die Level von Faktor A und $j = 1, \dots, J$ für die Level von Faktor B als Summe eines gruppenunspezifischen Erwartungswertes und den Effekten der Level von Faktor A und der Level von Faktor B modellieren.

Wir bezeichnen den gruppenunspezifischen Erwartungswertparameter mit μ_0 , den Effekt von Level i von Faktor A mit α_i und den Effekt von Level j von Faktor B mit β_j (β_j bezeichnet hier also nicht den j ten Eintrag des Betaparametervektors). Dann ergibt sich zum Beispiel für $I := J := 2$

$$\frac{\mu_{11} := \mu_0 + \alpha_1 + \beta_1}{\mu_{21} := \mu_0 + \alpha_2 + \beta_1} \quad \Bigg| \quad \frac{\mu_{12} := \mu_0 + \alpha_1 + \beta_2}{\mu_{22} := \mu_0 + \alpha_2 + \beta_2}$$

Wie im Falle der EVA ist diese Darstellung der Gruppenerwartungswerte μ_{ij} allerdings überparameterisiert. Um eine eindeutige Darstellung der μ_{ij} zu gewährleisten, bietet sich auch hier die Restriktion an, den Effekt des ersten Levels jedes Faktors als Null zu definieren

$$\alpha_1 := \beta_1 := 0. \tag{1}$$

und damit die Faktorlevelkombination A1B1 als Referenzgruppe zu etablieren. Es ergibt sich somit zum Beispiel für $I := J := 2$

$$\frac{\mu_{11} := \mu_0}{\mu_{21} := \mu_0 + \alpha_2} \quad \Bigg| \quad \frac{\mu_{12} := \mu_0 + \beta_2}{\mu_{22} := \mu_0 + \alpha_2 + \beta_2}$$

Auch bei dieser Effektdarstellung des Modells der additiven 2×2 ZVA mit Referenzgruppe ändern sich die Interpretation der Parameter μ_0, α_2, β_2 im Vergleich zum überparameterisierten Fall ohne Referenzgruppe: μ_0 entspricht dem Erwartungswert der Faktorlevelkombination A1B1, α_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A und β_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B.

Definition (Modell der additiven ZVA mit Referenzgruppe)

y_{ijk} mit $i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij}$ sei die Zufallsvariable, die den k ten Datenpunkt zum i ten Level von Faktor A und dem j ten Level von Faktor B in einem ZVA Anwendungsszenario modelliert. Dann hat das *Modell der additiven ZVA mit Referenzgruppe* die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (2)$$

und die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (3)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j \text{ f\"ur } i = 1, \dots, I, j = 1, \dots, J \text{ mit } \alpha_1 := \beta_1 := 0. \quad (4)$$

und $\sigma^2 > 0$.

Bemerkungen

- Das Modell der additiven ZVA modelliert ausschließlich Haupteffekte, keine Interaktionen.
- Wir verzichten auf die Angabe einer allgemeinen Designmatrixform dieses Modells.

Parameterbeispiele

(1) Es sei $\mu_0 := 1, \alpha_2 := 1, \beta_2 := 0$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 0 = \mathbf{1}$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 1 + 1 + 0 = \mathbf{2} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 1 + 1 + 0 = \mathbf{2}$$

⇒ Haupteffekt von Faktor A, kein Haupteffekt von Faktor B

(2) Es sei $\mu_0 := 1, \alpha_2 := 0, \beta_2 := 1$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 1 = \mathbf{2}$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 1 + 0 + 0 = \mathbf{1} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 1 + 0 + 1 = \mathbf{2}$$

⇒ Kein Haupteffekt von Faktor A, Haupteffekt von Faktor B

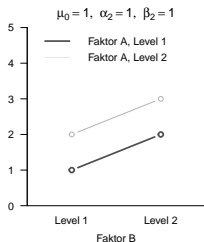
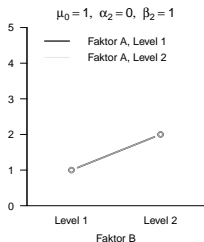
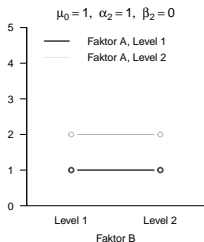
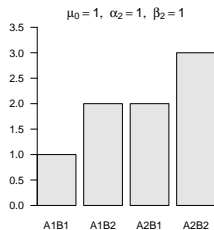
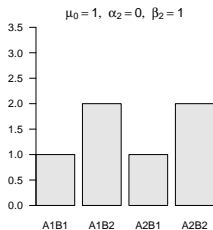
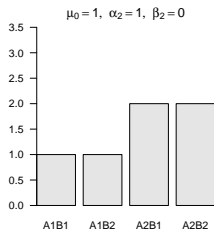
(3) Es sei $\mu_0 := 1, \alpha_2 := 1, \beta_2 := 1$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 1 = \mathbf{2}$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 1 + 1 + 0 = \mathbf{2} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 1 + 1 + 1 = \mathbf{3}$$

⇒ Haupteffekt von Faktor A, Haupteffekt von Faktor B

Parameterbeispiele



Theorem (Designmatrixform des Modells der 2 x 2 ZVA mit Referenzgruppe)

Gegeben sei die strukturelle Form des Modells der 2 x 2 ZVA mit Referenzgruppe. Dann hat dieses Modell die Designmatrixform

$$y \sim N(X\beta, \sigma^2 I_n), \quad (5)$$

wobei

$$y := \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times 3}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (6)$$

Bemerkung

- Das Theorem folgt direkt aus den Regeln der Matrixmultiplikation.

Modellformulierung

Modell der ZVA mit Interaktion

In der ZVA mit Interaktion möchte man die Gruppenerwartungswerte μ_{ij} mit $i = 1, \dots, I$ für die Level von Faktor A und $j = 1, \dots, J$ für die Level von Faktor B als Summe eines gruppenunspezifischen Erwartungswertes, der Effekte der Level von Faktor A und Faktor B und der Interaktion der Level der Faktoren modellieren.

Wir bezeichnen den gruppenunspezifischen Erwartungswertparameter mit μ_0 , den Effekt von Level i von Faktor A mit α_i , den Effekt von Level j von Faktor B mit β_j , und die Interaktion von Level i von Faktor A mit Level j von Faktor B mit γ_{ij} . Dann ergibt sich zum Beispiel für $I := J := 2$

$$\begin{array}{c|c} \mu_{11} := \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} & \mu_{12} := \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} \\ \mu_{21} := \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} \end{array}$$

Wie in der die additiven ZVA ist diese Darstellung der Gruppenerwartungswerte μ_{ij} multipel überparameterisiert. Um eine eindeutige Darstellung der μ_{ij} zu gewährleisten, bietet sich auch hier die Restriktion an, den Effekt des ersten Levels jedes Faktors und jeder Interaktion als Null zu definieren

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \dots, I, j = 1, \dots, J \quad (7)$$

und damit die Faktorlevelkombination A1B1 als Referenzgruppe zu etablieren. Es ergibt sich somit zum Beispiel für $I := J := 2$

$$\begin{array}{c|c} \mu_{11} := \mu_0 & \mu_{12} := \mu_0 + \beta_2 \\ \mu_{21} := \mu_0 + \alpha_2 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} \end{array}$$

Auch bei dieser Effektdarstellung des Modells der 2×2 ZVA mit Interaktion und Referenzgruppe ändern sich die Interpretation der Parameter $\mu_0, \alpha_2, \beta_2, \gamma_{22}$ im Vergleich zum überparameterisierten Fall ohne Referenzgruppe: μ_0 entspricht dem Erwartungswert der Faktorlevelkombination A1B1, α_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A, β_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B und γ_{22} der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B im Unterschiede zum Übergang von Level 1 zu Level 2 von Faktor A.

Definition (Modell der ZVA mit Interaktion und Referenzgruppe)

y_{ijk} mit $i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij}$ sei die Zufallsvariable, die den k ten Datenpunkt zum i ten Level von Faktor A und dem j ten Level von Faktor B in einem ZVA Anwendungsszenario modelliert. Dann hat das Modell der ZVA mit Interaktion und Referenzgruppe die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (8)$$

und die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (9)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j + \gamma_{ij} \quad (10)$$

sowie

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ f\"ur } i = 1, \dots, I, j = 1, \dots, J \quad (11)$$

und $\sigma^2 > 0$.

Bemerkungen

- Wir verzichten auf die Angabe einer allgemeinen Designmatrixform dieses Modells.

Parameterbeispiele

(1) Es sei $\mu_0 := 1, \alpha_2 := 0, \beta_2 := 0, \gamma_{22} = 2$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 0 + 0 = \mathbf{1}$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 0 + 0 = \mathbf{1} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 0 + 0 + 2 = \mathbf{3}$$

(2) Es sei $\mu_0 := 1, \alpha_2 := 1, \beta_2 := 1, \gamma_{22} = -2$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 1 + 0 = \mathbf{2}$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 1 + 0 = \mathbf{2} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 1 + 1 - 2 = \mathbf{1}$$

(3) Es sei $\mu_0 := 1, \alpha_2 := 1, \beta_2 := 0, \gamma_{22} = 1$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 0 + 0 = \mathbf{1}$$

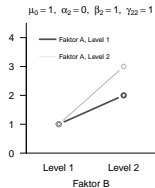
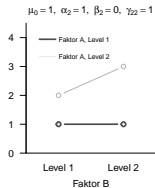
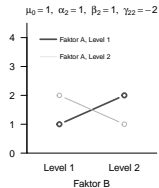
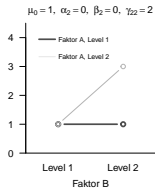
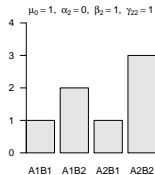
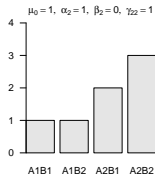
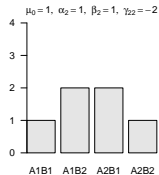
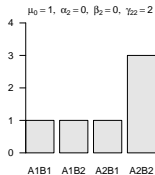
$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 1 + 0 + 0 = \mathbf{2} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 1 + 0 + 1 = \mathbf{3}$$

(4) Es sei $\mu_0 := 1, \alpha_2 := 0, \beta_2 := 1, \gamma_{22} = 1$. Dann gilt:

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = \mathbf{1} \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 1 + 0 = \mathbf{2}$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 0 + 0 = \mathbf{1} \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 0 + 1 + 1 = \mathbf{3}$$

Parameterbeispiele



Theorem (Designmatrixform des Modells der 2 x 2 ZVA mit Interaktion und Referenzgruppe)

Gegeben sei die strukturelle Form eines 2 x 2 ZVA Modells mit Interaktion und Referenzgruppe und es sei

$$n := \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \quad (12)$$

die Gesamtanzahl an Datenvariablen. Dann hat dieses Modell die Designmatrixform

$$y \sim N(X\beta, \sigma^2 I_n), \text{ mit} \quad (13)$$

$$y := \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times 4}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (14)$$

Modellformulierung

Beispiel

Es seien

$$I := 2, J := 2 \text{ und } n_{ij} := 4 \text{ für } i = 1, 2, j = 1, 2, \text{ also } n = 16. \quad (15)$$

Dann gilt

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{16}, \sigma^2 I_{16}) \quad (16)$$

mit

$$y := \begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{124} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{224} \end{pmatrix}, X := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{16 \times 4}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (17)$$

Modellformulierung

Beispiel

```
# Modellformulierung
library(MASS)
I      = 2
J      = 2
n_ij   = 4
n      = I*J*n_ij
p      = 1 + (I-1)+(J-1)+(I*J-3)
D      = matrix(c(1,0,0,0,
                 1,0,1,0,
                 1,1,0,0,
                 1,1,1,1),
               nrow = p,
               byrow = TRUE)

C      = matrix(rep(1,n_ij),nrow = n_ij)
X      = kronecker(D,C)
I_n    = diag(n)
beta   = matrix(c(1,1,1,1), nrow = p)
sigsqr = 10

# Multivariate Normalverteilung
# Anzahl Level Faktor A
# Anzahl Level Faktor B
# Anzahl von Datenpunkten der i,jten Gruppe
# Anzahl Datenpunkte
# Anzahl Parameter
# Prototypische Designmatrix für balancierte Designs

# Prototypischer Zellenvektor für balancierte Designs
# Kroneckerprodukt Designmatrix Erzeugung für balancierte Designs
# n x n Einheitsmatrix
# \beta = (\mu_0, \alpha_2, \alpha_3, \alpha_4)
# \sigma^2

# Datenrealisierung
y      = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
print(X)
```

```
>      [,1] [,2] [,3] [,4]
> [1,]  1   0   0   0
> [2,]  1   0   0   0
> [3,]  1   0   0   0
> [4,]  1   0   0   0
> [5,]  1   0   1   0
> [6,]  1   0   1   0
> [7,]  1   0   1   0
> [8,]  1   0   1   0
> [9,]  1   1   0   0
> [10,] 1   1   0   0
> [11,] 1   1   0   0
> [12,] 1   1   0   0
> [13,] 1   1   1   1
> [14,] 1   1   1   1
> [15,] 1   1   1   1
> [16,] 1   1   1   1
```

Beispieldatensatzerzeugung

```
# Datensimulation
library(MASS)
set.seed(1)
I      = 2
J      = 2
n_ij   = 20
n      = I*J*n_ij
p      = 1 + (I-1)+(J-1)+(I*J-3)
D      = matrix(c(1,0,0,0,
                 1,0,1,0,
                 1,1,0,0,
                 1,1,1,1),
               nrow = p,
               byrow = TRUE)

C      = matrix(rep(1,n_ij),nrow = n_ij)
X      = kronecker(D,C)
I_n    = diag(n)
beta   = matrix(c(7,-3,0,-2), nrow = p)
sigsqr = 4
y      = mvrnorm(1, X %*% beta, sigsqr*I_n)

# Multivariate Normalverteilung
# reproduzierbare Daten
# Anzahl Level Faktor A (Therapie)
# Anzahl Level Faktor B (Alter)
# Anzahl von Datenpunkten der i,jten Gruppe
# Anzahl Datenpunkte
# Anzahl Parameter
# Prototypische Designmatrix für balancierte Designs

C      = matrix(rep(1,n_ij),nrow = n_ij)
X      = kronecker(D,C)
I_n    = diag(n)
beta   = matrix(c(7,-3,0,-2), nrow = p)
sigsqr = 4
y      = mvrnorm(1, X %*% beta, sigsqr*I_n)

# Prototypischer Zellenvektor für balancierte Designs
# Kroneckerprodukt für balancierte Designs
# n x n Einheitsmatrix
# beta = (\mu_0, \alpha_2, \beta_2, \gamma_22)
# \sigma^2
# eine Realisierung eines n-dimensionalen ZVs

# Dataframeformatierung
library(writexl)
D      = data.frame("ID" = 1:n)
D$Therapy = c(rep("F2F", J*n_ij), rep("ONL", J*n_ij))
D$Age    = rep(c(rep("YA", n_ij), rep("OA", n_ij)),I)
D$BDI    = y

# Excel Output
# Dataframe Initialisierung und ID Variable
# Therapiebedingung
# Alter
# PrePost-BDI Differenzwerte

# Datenspeicherung
write_xlsx(D, file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.xlsx"))
write_csv(D, file = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv"))
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Theorem (Betaparameterschätzung im additiven 2×2 ZVA Modell mit Referenzgruppe)

Gegeben sei die Designmatrixform eines balancierten additiven 2×2 ZVA Modells mit Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4}\bar{y}_{11} + \frac{1}{4}(\bar{y}_{12} + \bar{y}_{21}) - \frac{1}{4}\bar{y}_{22} \\ \frac{1}{2}(\bar{y}_{21} + \bar{y}_{22}) - \frac{1}{2}(\bar{y}_{11} + \bar{y}_{12}) \\ \frac{1}{2}(\bar{y}_{12} + \bar{y}_{22}) - \frac{1}{2}(\bar{y}_{11} + \bar{y}_{21}) \end{pmatrix}, \quad (18)$$

wobei

$$\bar{y}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \text{ für } 1 \leq i, j \leq 2 \quad (19)$$

das Stichprobenmittel der i, j ten Gruppe des 2×2 ZVA Designs bezeichnet.

Beweis

Wir bestimmen zunächst $X^T y$, $X^T X$ und $(X^T X)^{-1}$ bei konstantem n_{ij} für $1 \leq i, j \leq 2$.

Modellschätzung

Beweis (fortgeführt)

$$X^T y = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \\ \vdots & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ y_{11n_{11}} & & & & & & & & & & & \\ y_{121} & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ y_{12n_{12}} & & & & & & & & & & & \\ y_{211} & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ y_{21n_{21}} & & & & & & & & & & & \\ y_{221} & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ \vdots & & & & & & & & & & & \\ y_{22n_{22}} & & & & & & & & & & & \end{pmatrix} \quad (20)$$
$$= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \end{pmatrix}.$$

Beweis (fortgeführt)

$$X^T X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix} \quad (21)$$
$$= n_{ij} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$$

Beweis (fortgeführt)

Ohne Beweis halten wir weiterhin fest, dass

$$(X^T X)^{-1} = n_{ij} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}^{-1} = \frac{1}{n_{ij}} \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \quad (22)$$

Es ergibt sich also

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \frac{1}{n_{ij}} \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \end{pmatrix}. \quad (23)$$

Beweis (fortgeführt)

Damit ergibt sich dann

$$\begin{aligned}\hat{\mu}_0 &= \frac{1}{n_{ij}} \left(\frac{3}{4} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} - \frac{1}{2} \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} - \frac{1}{2} \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \right) \\ &= \frac{1}{n_{ij}} \left(\frac{3}{4} \sum_{k=1}^{n_{11}} y_{11k} + \frac{3}{4} \sum_{k=1}^{n_{12}} y_{12k} + \frac{3}{4} \sum_{k=1}^{n_{21}} y_{21k} + \frac{3}{4} \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &+ \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{k=1}^{n_{21}} y_{21k} - \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} - \frac{1}{2} \sum_{k=1}^{n_{12}} y_{12k} - \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{ij}} \left(\frac{3}{4} \sum_{k=1}^{n_{11}} y_{11k} + \frac{1}{4} \sum_{k=1}^{n_{12}} y_{12k} + \frac{1}{4} \sum_{k=1}^{n_{21}} y_{21k} - \frac{1}{4} \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{3}{4} \bar{y}_{11} + \frac{1}{4} (\bar{y}_{12} + \bar{y}_{21}) - \frac{1}{4} \bar{y}_{22}\end{aligned} \tag{24}$$

Beweis (fortgeführt)

sowie

$$\begin{aligned}\hat{\alpha}_2 &= \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} + \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \right) \\ &= \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{k=1}^{n_{11}} y_{11k} - \frac{1}{2} \sum_{k=1}^{n_{12}} y_{12k} - \frac{1}{2} \sum_{k=1}^{n_{21}} y_{21k} - \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} + \sum_{k=1}^{n_{21}} y_{21k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{k=1}^{n_{11}} y_{11k} - \frac{1}{2} \sum_{k=1}^{n_{12}} y_{12k} + \frac{1}{2} \sum_{k=1}^{n_{21}} y_{21k} + \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{2} (\bar{y}_{21} + \bar{y}_{22}) - \frac{1}{2} (\bar{y}_{11} + \bar{y}_{12})\end{aligned}\tag{25}$$

und analog für $\hat{\beta}_2$.

Modellschätzung

Beispiel

```
# Datenreformatierung
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE) # Datensatz
A1B1 = D$BDI[D$Therapy == "F2F" & D$Age == "YA" ] # Face-to-face, younger adults
A1B2 = D$BDI[D$Therapy == "F2F" & D$Age == "OA" ] # Face-to-face, younger adults
A2B1 = D$BDI[D$Therapy == "ONL" & D$Age == "YA" ] # Online, older adults
A2B2 = D$BDI[D$Therapy == "ONL" & D$Age == "OA" ] # Online, older adults

# Datenmatrix für Gruppenmittelwerte
n_ij = length(A1B1) # Anzahl Datenpunkte pro Gruppe
Y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n_ij) # Datenmatrix
bar_y = colMeans(Y) # Zellenmittelwerte

# Modellschätzung
I = 2 # Anzahl Level Faktor A (Therapie)
J = 2 # Anzahl Level Faktor B (Alter)
n = I*J*n_ij # Anzahl Datenpunkte
p = 1 + (I-1)+(J-1)+(I*J-3) # Anzahl Parameter
D = matrix(c(1,0,0, # Prototypische Designmatrix für balancierte Designs
            1,0,1,
            1,1,0,
            1,1,1), nrow = p, byrow = TRUE)

C = matrix(rep(1,n_ij),nrow = n_ij) # Prototypischer Zellenvektor für balancierte Designs
X = kronecker(D,C) # Kroneckerprodukt Designmatrix
y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
cat("hat{beta} : ", beta_hat, # Ausgabe
    "\nhat{sigsqr} : ", sigsq_hat,
    "\ny_11,y_12,y_21,y_22 : ", bar_y,
    "\n3/4y_11 + 1/4(y_12 + y_21) - 1/4y_22 : ", (3/4)*bar_y[1]+(1/4)*(bar_y[2]+bar_y[3])-(1/4)*bar_y[4],
    "\n1/2(y_21 + y_22) - 1/2(y_11 + y_12) : ", (1/2)*(bar_y[3]+bar_y[4])-(1/2)*(bar_y[1] + bar_y[2]),
    "\n1/2(y_12 + y_22) - 1/2(y_11 + y_21) : ", (1/2)*(bar_y[2]+bar_y[4])-(1/2)*(bar_y[1] + bar_y[3]))

> hat{beta} : 7.62 -4.06 -0.766
> hat{sigsqr} : 3.54
> y_11,y_12,y_21,y_22 : 7.2 7.28 3.99 2.38
> 3/4y_11 + 1/4(y_12 + y_21) - 1/4y_22 : 7.62
> 1/2(y_21 + y_22) - 1/2(y_11 + y_12) : -4.06
> 1/2(y_12 + y_22) - 1/2(y_11 + y_21) : -0.766
```

Theorem (Betaparameterschätzung im 2×2 ZVA Modell mit Interaktion und Referenzgruppe)

Gegeben sei die Designmatrixform eines balancierten 2×2 ZVA Modells mit Interaktion und Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \begin{pmatrix} \bar{y}_{11} \\ \bar{y}_{21} - \bar{y}_{11} \\ \bar{y}_{12} - \bar{y}_{11} \\ \bar{y}_{11} + \bar{y}_{22} - \bar{y}_{12} - \bar{y}_{21} \end{pmatrix}, \quad (26)$$

wobei

$$\bar{y}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \quad \text{für } 1 \leq i, j \leq 2 \quad (27)$$

das Stichprobenmittel der i, j ten Gruppe des 2×2 ZVA Designs bezeichnet.

Beweis

Wir bestimmen zunächst $X^T y$, $X^T X$ und $(X^T X)^{-1}$ bei konstantem n_{ij} für $1 \leq i, j \leq 2$.

Modellschätzung

Beweis (fortgeführt)

$$X^T y = \begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_{111} \\ \vdots \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ \vdots \\ y_{22n_{22}} \end{pmatrix} \quad (28)$$
$$= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \\ \sum_{k=1}^{n_{22}} y_{22k} \end{pmatrix}.$$

Beweis (fortgeführt)

$$X^T X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad (29)$$
$$= n_{ij} \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Beweis (fortgeführt)

Ohne Beweis halten wir weiterhin fest, dass

$$(X^T X)^{-1} = n_{ij} \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}^{-1} = \frac{1}{n_{ij}} \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 4 \end{pmatrix}. \quad (30)$$

Es ergibt sich also

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \frac{1}{n_{ij}} \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 4 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \\ \sum_{k=1}^{n_{22}} y_{22k} \end{pmatrix}. \quad (31)$$

Beweis (fortgeführt)

Damit ergibt sich dann

$$\begin{aligned}\hat{\mu}_0 &= \frac{1}{n_{ij}} \left(\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} - \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} - \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{11}} y_{11k} + \sum_{k=1}^{n_{12}} y_{12k} + \sum_{k=1}^{n_{21}} y_{21k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &+ \frac{1}{n_{ij}} \left(- \sum_{k=1}^{n_{21}} y_{21k} - \sum_{k=1}^{n_{22}} y_{22k} - \sum_{k=1}^{n_{12}} y_{12k} - \sum_{k=1}^{n_{22}} y_{22k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} y_{11k} \\ &= \bar{y}_{11}\end{aligned} \tag{32}$$

Beweis (fortgeführt)

sowie

$$\begin{aligned}\hat{\alpha}_2 &= \frac{1}{n_{ij}} \left(- \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} + 2 \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} + 1 \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} - 2 \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{ij}} \left(- \sum_{k=1}^{n_{11}} y_{11k} - \sum_{k=1}^{n_{12}} y_{12k} - \sum_{k=1}^{n_{21}} y_{21k} - \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &+ \frac{1}{n_{ij}} \left(2 \sum_{k=1}^{n_{21}} y_{21k} + 2 \sum_{k=1}^{n_{22}} y_{22k} + \sum_{k=1}^{n_{12}} y_{12k} + \sum_{k=1}^{n_{22}} y_{22k} - 2 \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{21}} y_{21k} - \sum_{k=1}^{n_{11}} y_{11k} \right) \\ &= \bar{y}_{21} - \bar{y}_{11}\end{aligned} \tag{33}$$

und analog für $\hat{\beta}_2$.

Beweis (fortgeführt)

Schließlich ergibt sich

$$\begin{aligned}\hat{\gamma}_{22} &= \frac{1}{n_{ij}} \left(\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} - 2 \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} - 2 \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} + 4 \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{11}} y_{11k} + \sum_{k=1}^{n_{12}} y_{12k} + \sum_{k=1}^{n_{21}} y_{21k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\ &+ \frac{1}{n_{ij}} \left(-2 \sum_{k=1}^{n_{21}} y_{21k} - 2 \sum_{k=1}^{n_{22}} y_{22k} - 2 \sum_{k=1}^{n_{12}} y_{12k} - 2 \sum_{k=1}^{n_{22}} y_{22k} + 4 \sum_{k=1}^{n_{22}} y_{22k} \right) \quad (34) \\ &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{11}} y_{11k} + \sum_{k=1}^{n_{22}} y_{22k} - \sum_{k=1}^{n_{12}} y_{12k} - \sum_{k=1}^{n_{21}} y_{21k} \right) \\ &= \bar{y}_{11} + \bar{y}_{22} - \bar{y}_{12} - \bar{y}_{21}.\end{aligned}$$

Modellschätzung

Beispiel

```
# Datenreformatierung
fname      = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D          = read.table(fname, sep = ",", header = TRUE)
A1B1      = D$BDI[D$Therapy == "F2F" & D$Age == "YA" ]
A1B2      = D$BDI[D$Therapy == "F2F" & D$Age == "OA" ]
A2B1      = D$BDI[D$Therapy == "ONL" & D$Age == "YA" ]
A2B2      = D$BDI[D$Therapy == "ONL" & D$Age == "OA" ]

# Datenmatrix für Gruppenmittelwerte
n_ij       = length(A1B1)
Y          = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n_ij)
bar_y     = colMeans(Y)

# Modellschätzung
I          = 2
J          = 2
n          = I*J*n_ij
p          = 1 + (I-1)+(J-1)+(I*J-3)
D          = matrix(c(1,0,0,0,
                    1,0,1,0,
                    1,1,0,0,
                    1,1,1,1), nrow = p, byrow = TRUE)

C          = matrix(rep(1,n_ij),nrow = n_ij)
X          = kronecker(D,C)
y          = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n)
beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y
eps_hat   = y - X %*% beta_hat
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)
cat("hat{beta}          : ", beta_hat,
    "\nhat{sigsqr}       : ", sigsqr_hat,
    "\ny_11,y_12,y_21,y_22 : ", bar_y,
    "\ny_11              : ", bar_y[1],
    "\ny_21 - y_11       : ", bar_y[3]-bar_y[1],
    "\ny_12 - y_11       : ", bar_y[2]-bar_y[1],
    "\ny_11 + y_22 - y_12 + y_21 : ", bar_y[1]+bar_y[4]-bar_y[3]-bar_y[2])

> hat{beta}          : 7.2 -3.22 0.0741 -1.68
> hat{sigsqr}       : 3.35
> y_11,y_12,y_21,y_22 : 7.2 7.28 3.99 2.38
> y_11              : 7.2
> y_21 - y_11       : -3.22
> y_12 - y_11       : 0.0741
> y_11 + y_22 - y_12 + y_21 : -1.68
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Überblick

Wie bei der EVA kann auch bei der ZVA eine Modellevaluationstheorie mithilfe einer Quadratsummenzerlegung entwickelt werden (vgl. Inferenzstatistik SoSe 2021). Mit zunehmender Designkomplexität wird eine solche Theorie allerdings zunehmend unübersichtlich. Prinzipiell können alle Quadratsummenzerlegung-basierten F-Statistiken auf Likelihood-basierte Modellvergleich zurückgeführt werden. Allerdings fehlt dazu in diesem Kurs bisher eine allgemeine Kontrasttheorie. Wir beschränken und im Folgenden aufgrund dieser Tatsachen deshalb auf die

- (1) Evaluation der Haupteffekte im additiven Modell der 2×2 ZVA mit Referenzgruppe und die
- (2) Evaluation der Interaktion im Modell der 2×2 ZVA mit Interaktion und Referenzgruppe

Zukünftige Iterationen dieses Kurses werden eine allgemeine Kontrasttheorie und ihren Bezug zu Quadratsummenzerlegungen und Modellvergleichen beinhalten, siehe dazu zum Beispiel Poline (2012) *Neuroimage* 15;62(2):871-80 und Christensen (2011) *Plane Answers to Complex Questions: The Theory of General Linear Models*.

Evaluation der Haupteffekte im additiven Modell der 2×2 ZVA mit Referenzgruppe

- (1) Statistische Modelle und Teststatistiken
- (2) Testhypothesen und Tests
- (3) Testumfangkontrollen und p-Werte

Evaluation der Interaktion im Modell der 2×2 ZVA mit Interaktion und Referenzgruppe

- (1) Statistisches Modell und Teststatistik
- (2) Testhypothese und Test
- (3) Testumfangkontrolle und p-Wert

Theorem (Teststatistiken für Haupteffekte)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (35)$$

die Designmatrixform des additiven Modells der 2×2 ZVA mit Referenzgruppe, wobei die $n \times 1$ Spalten von X bezeichnet seien durch

$$X := \begin{pmatrix} X_{\mu_0} & X_{\alpha_2} & X_{\beta_2} \end{pmatrix} \in \mathbb{R}^{n \times 3} \quad (36)$$

Dann gelten

(A) Eine *F-Teststatistik für den Haupteffekt von Faktor A* ist die F-Statistik des ALMs mit

$$X_A := \begin{pmatrix} X_{\mu_0} & X_{\beta_2} & X_{\alpha_2} \end{pmatrix} \in \mathbb{R}^{n \times 3}, \beta_A := \begin{pmatrix} \mu_0 \\ \beta_2 \\ \alpha_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } p_1 := 2, p_2 := 1. \quad (37)$$

(B) Eine *F-Teststatistik für den Haupteffekt von Faktor B* ist die F-Statistik des ALMs mit

$$X_B := \begin{pmatrix} X_{\mu_0} & X_{\alpha_2} & X_{\beta_2} \end{pmatrix} \in \mathbb{R}^{n \times 3}, \beta_B := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } p_1 := 2, p_2 := 1. \quad (38)$$

Definition (Testhypothesen und Tests)

Gegeben sei das Modell der additiven 2×2 ZVA und die F-Teststatistiken für die Haupteffekte von Faktor A und B seien mit F_A , F_B und $F_{A \times B}$ bezeichnet und wie oben definiert. Dann gelten folgende Definitionen:

(A) Der kritische Wert-basierte Test

$$\phi_A(y) := 1_{\{F_A \geq k\}} \text{ mit Nullhypothese } H_0^A : \alpha_2 = 0 \quad (39)$$

definiert den *F-Test des Haupteffekts von Faktor A* im Modell mit X_A, β_A .

(B) Der kritische Wert-basierte Test

$$\phi_B(y) := 1_{\{F_B \geq k\}} \text{ mit Nullhypothese } H_0^B : \beta_2 = 0 \quad (40)$$

definiert den *F-Test des Haupteffekts von Faktor B* im Modell mit X_B, β_B .

Theorem (Testumfangkontrolle und p-Werte)

Mit obigen Definition und der KVF φ der f -Verteilung gelten:

(A) ϕ_A ist ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert ist definiert durch

$$k_{\alpha_0}^A := \varphi^{-1}(1 - \alpha_0; 1, n - 3) \quad (41)$$

und der zu einem beobachteten Wert f_A von F_A assoziierte p-Wert ist gegeben durch

$$p_A\text{-Wert} := 1 - \varphi(1 - \alpha_0; 1, n - 3). \quad (42)$$

(B) ϕ_B ist ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert ist definiert durch

$$k_{\alpha_0}^B := \varphi^{-1}(1 - \alpha_0; 1, n - 3) \quad (43)$$

und der zu einem beobachteten Wert f_B von F_B assoziierte p-Wert ist gegeben durch

$$p_B\text{-Wert} := 1 - \varphi(1 - \alpha_0; 1, n - 3). \quad (44)$$

Bemerkung

- Wir verzichten auf einen Beweis.

Modellevaluation | Haupteffekte im additiven Modell der 2 x 2 ZVA mit Referenzgruppe

```

# Modellevaluation
I      = 2                                # Anzahl Level Faktor A (Therapie)
J      = 2                                # Anzahl Level Faktor B (Alter)
n_ij   = length(A1B1)                    # balanciertes ANOVA Design
n      = I*J*n_ij                        # Anzahl Datenpunkte
p      = 3                                # Anzahl Parameter vollständiges Modell
y      = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
D      = matrix(c(1,0,0,1,0,1,1,1,0,1,1,1), # Prototypische Designmatrix
               nrow = I*J,byrow=TRUE)
C      = matrix(rep(1,n_ij),nrow = n_ij)  # Prototypischer Zellenvektor für balancierte Designs
X      = kronecker(D,C)                  # ZVA Kroneckerprodukt Designmatrix
XH     = list(X,c(1,3,2), X)            # Modellvarianten
alpha_0 = 0.05                          # Signifikanzlevel
Eff    = rep(NA,n,2)                   # F-Teststatistik Arrayinitialisierung
k_alpha_0 = rep(NA,n,2)                # Kritischer Wert Arrayinitialisierung
phi    = rep(NA,n,2)                   # Testwert Arrayinitialisierung
p_vals = rep(NA,n,2)                   # p-Wert Arrayinitialisierung
for(i in 1:2){
  X      = XH[[i]]                      # Designmatrix vollständiges Modell
  X_1    = X[,-3]                       # Designmatrix reduziertes Modell
  p      = ncol(X)                       # Anzahl Parameter vollständiges Modell
  p_1    = ncol(X_1)                     # Anzahl Parameter reduziertes Modell
  p_2    = p - p_1                       # Anzahl zusätzlicher Parameter im vollst. Modell
  beta_hat_1 = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y # Betaparameterschätzer reduziertes Modell
  beta_hat   = solve(t(X) %*%X )%*%t(X) %*%y   # Betaparameterschätzer vollständiges Modell
  eps_hat_1  = y-X_1)%*%beta_hat_1           # Residuenvektor reduziertes Modell
  eps_hat    = y - X)%*%beta_hat            # Residuenvektor vollständiges Modell
  eh1_eh1    = t(eps_hat_1) %*% eps_hat_1    # RQS reduziertes Modell
  eh_eh      = t(eps_hat) %*% eps_hat        # RQS vollständiges Modell
  sigsqqr_hat = eh_eh/(n-p)                 # Varianzparameterschätzer vollst. Modell
  Eff[i]     = ((eh1_eh1-eh_eh)/p_2)/sigsqqr_hat # F-Statistik
  k_alpha_0[i] = qf(1-alpha_0, p_2, n-p)     # kritischer Wert
  if(Eff[i] >= k_alpha_0[i]){ phi[i] = 1 }   # H_0 Ablehnen
  else { phi[i] = 0 }                       # H_A Annehmen
  p_vals[i]  = 1 - pf(Eff[i], p_2,n-p)      # p-Wert
}
data.frame("f"= Eff,"k" = k_alpha_0,"phi"= phi, "p-Wert" = p_vals, row.names = c("Therapy", "Age"))

>
>      f      k phi      p.Wert
> Therapy 94.25 3.97 1 5.22e-15
> Age      3.36 3.97 0 7.07e-02

```

Haupteffekte der additiven 2 x 2 ZVA mit R's aov() Funktion

```
# Dateneinlesen
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# R's aov Funktion
res.aov = aov(BDI ~ Therapy + Age, data = D)
summary(res.aov)                                     # Modellformulierung und Modellschätzung
                                                    # Modellevaluation

>
> Df Sum Sq Mean Sq F value Pr(>F)
> Therapy 1 329 329 94.25 5.2e-15 ***
> Age 1 12 12 3.36 0.071 .
> Residuals 77 269 3
> ---
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Haupteffekte der additiven 2 x 2 ZVA mit R's lm() und anova() Funktionen

```
# Dateneinlesen
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# R's lm (linear model) und anova() Funktionen
glm = lm(BDI ~ Therapy + Age, data = D)
anova(glm)                                           # Modellformulierung und Modellschätzung
                                                    # Modellevaluation

> Analysis of Variance Table
>
> Response: BDI
> Df Sum Sq Mean Sq F value Pr(>F)
> Therapy 1 329 329 94.25 5.2e-15 ***
> Age 1 12 12 3.36 0.071 .
> Residuals 77 269 3
> ---
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Evaluation der Haupteffekte im additiven Modell der 2×2 ZVA mit Referenzgruppe

- (1) Statistische Modelle und Teststatistiken
- (2) Testhypothesen und Tests
- (3) Testumfangkontrollen und p-Werte

Evaluation der Interaktion im Modell der 2×2 ZVA mit Interaktion und Referenzgruppe

- (1) Statistisches Modell und Teststatistik
- (2) Testhypothe und Test
- (3) Testumfangkontrolle und p-Wert

Theorem (Teststatistik für die Interaktion)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (45)$$

die Designmatrixform des 2×2 ZVA Modells mit Interaktion und Referenzgruppe. Dann ist die F-Statistik mit $p_1 := 3$ und $p_2 := 1$ eine F-Teststatistik für die Interaktion von Faktor A und Faktor B.

Definition (Testhypothese und Test)

Die F-Teststatistik für die Interaktion von Faktor A und Faktor B sei mit $F_{A \times B}$ bezeichnet und wie oben definiert. Dann definiert der kritische Wert-basierte Test

$$\phi_{A \times B}(y) := 1_{\{F_{A \times B} \geq k\}} \text{ mit Nullhypothese } H_0^{A \times B} : \gamma_{22} = 0 \quad (46)$$

den *F-Test der Interaktion von Faktor A und Faktor B*.

Theorem (Testumfangkontrolle und p-Wert)

Mit obigen Definition und der KVF φ der f -Verteilung gilt, dass $\phi_{A \times B}$ ist ein Level- α_0 -Test mit Testumfang α_0 ist, wenn der kritische Wert ist definiert durch

$$k_{\alpha_0}^{A \times B} := \varphi^{-1}(1 - \alpha_0; 1, n - 4) \quad (47)$$

und der zu einem beobachteten Wert $f_{A \times B}$ von $F_{A \times B}$ assoziierte p-Wert ist gegeben durch

$$p_{A \times B}\text{-Wert} := 1 - \varphi(1 - \alpha_0; 1, n - 4). \quad (48)$$


```

# Modellevaluation
I      = 2                                # Anzahl Level Faktor A (Therapie)
J      = 2                                # Anzahl Level Faktor B (Alter)
n_ij   = length(A1B1)                    # balanciertes ANOVA Design
n      = I*J*n_ij                         # Anzahl Datenpunkte
p      = 4                                # Anzahl Parameter vollständiges Modell
y      = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
D      = matrix(c(1,0,0,0,1,0,1,0, 1,1,0,0,1,1,1,1), # Prototypische Designmatrix
               nrow = I*J,byrow=TRUE)
C      = matrix(rep(1,n_ij),nrow = n_ij)   # Prototypischer Zellenvektor für balancierte Designs
X      = kronecker(D,C)                   # ZVA Kroneckerprodukt Designmatrix
XH     = list(X,c(1,3,2)), X              # Modellvarianten
alpha_0 = 0.05                           # Signifikanzlevel
X      = XH[[i]]                          # Designmatrix vollständiges Modell
X_1    = X[,-4]                           # Designmatrix reduziertes Modell
p      = ncol(X)                          # Anzahl Parameter vollständiges Modell
p_1    = ncol(X_1)                        # Anzahl Parameter reduziertes Modell
p_2    = p - p_1                          # Anzahl zusätzlicher Parameter im vollst. Modell
beta_hat_1 = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y # Betaparameterschätzer reduziertes Modell
beta_hat  = solve(t(X)%*%X)%*%t(X)%*%y     # Betaparameterschätzer vollständiges Modell
eps_hat_1 = y-X_1%*%beta_hat_1            # Residuenvektor reduziertes Modell
eps_hat   = y - X%*%beta_hat             # Residuenvektor vollständiges Modell
eh1_eh1   = t(eps_hat_1)%*%eps_hat_1      # RQS reduziertes Modell
eh_eh     = t(eps_hat)%*%eps_hat          # RQS vollständiges Modell
sigsqr_hat = eh_eh/(n-p)                  # Varianzparameterschätzer vollst. Modell
f         = ((eh1_eh1-eh_eh)/p_2)/sigsqr_hat # F-Statistik
k_alpha_0 = qf(1-alpha_0, p_2, n-p)        # kritischer Wert
if(f >= k_alpha_0){phi = 1} else {phi = 0} # Test
p_val     = 1 - pf(f, p_2,n-p)            # p-Wert
data.frame("f"= f,"k" = k_alpha_0," phi"= phi, "p-Wert" = p_val, row.names = c("Therapy x Age"))

>
> f k X.phi p.Wert
> Therapy x Age 4.21 3.97 1 0.0436

```

Interaktion im Modell der 2 x 2 ZVA mit Interaktion mit R's aov() Funktion

```
# Dateneinlesen
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# R's aov Funktion
res.aov = aov(BDI ~ Therapy + Age + Age:Therapy, data = D) # Modellformulierung und Modellschätzung
summary(res.aov) # Modellevaluation

>
> Df Sum Sq Mean Sq F value Pr(>F)
> Therapy 1 329 329 98.18 2.5e-15 ***
> Age 1 12 12 3.50 0.065 .
> Therapy:Age 1 14 14 4.21 0.044 *
> Residuals 76 255 3
> ---
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaktion im Modell der 2 x 2 ZVA mit Interaktion mit R's lm() und anova() Funktionen

```
# Dateneinlesen
fname = file.path(getwd(), "11_Daten", "11_Zweifaktorielle_Varianzanalyse_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE)

# R's lm (linear model) und anova() Funktionen
glm = lm(BDI ~ Therapy + Age + Therapy:Age, data = D) # Modellformulierung und Modellschätzung
anova(glm) # Modellevaluation

> Analysis of Variance Table
>
> Response: BDI
> Df Sum Sq Mean Sq F value Pr(>F)
> Therapy 1 329 329 98.18 2.5e-15 ***
> Age 1 12 12 3.50 0.065 .
> Therapy:Age 1 14 14 4.21 0.044 *
> Residuals 76 255 3
> ---
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario der zweifaktoriellen Varianzanalyse (ZVA).
2. Aus wie vielen Datenpunkten besteht ein Datensatz eines 3×4 ZVA Designs mit 10 Datenpunkten pro Zelle?
3. Erläutern Sie die intuitive Bedeutung eines Haupteffektes in einem ZVA Design.
4. Erläutern Sie die intuitive Bedeutung einer Interaktion in einem ZVA Design.
5. Geben Sie die Definition des additiven Modells der ZVA mit Referenzgruppe (RG) wieder.
6. Erläutern Sie die Bedeutung der Parameter μ_0 , α_2 und β_2 im additiven Modell der ZVA mit RG.
7. Bestimmen Sie μ_{ij} für $\mu_0 := 2$, $\alpha_2 = -1$ und $\beta_2 := 3$ im additiven Modell der ZVA mit RG.
8. Geben Sie die Designmatrixform des Modells einer additiven 2×2 ZVA mit RG für $n_{ij} := 1$ an.
9. Geben Sie die Designmatrixform des Modells einer additiven 2×2 ZVA mit RG für $n_{ij} := 3$ an.
10. Geben Sie die Definition des Modells der ZVA mit Interaktion und RG wieder.
11. Erläutern Sie die Bedeutung der Parameter μ_0 , α_2 , β_2 und γ_{22} im Modell der ZVA mit Interaktion und RG.
12. Geben Sie die Designmatrixform des Modells einer 2×2 ZVA mit Interaktion und RG für $n_{ij} := 1$ an.
13. Geben Sie die Designmatrixform des Modells einer 2×2 ZVA mit Interaktion und RG für $n_{ij} := 3$ an.
14. Geben Sie das Theorem zur Betaparameterschätzung im additiven 2×2 ZVA Modell mit RG wieder.
15. Geben Sie das Theorem zur Betaparameterschätzung im 2×2 ZVA Modell mit Interaktion und RG wieder.

16. Erzeugen Sie einen Beispieldatensatz des Anwendungsbeispiels einer 2×2 ZVA mit den Faktoren Therapie (F2F, ONL) und Alter (YA,OA) basierend auf dem Modell der ZVA mit Interaktion und Referenzgruppe für von Ihnen selbst gewählte Parameterwerte μ_0 , α_2 , β_2 , γ_{22} und σ^2 . Visualisieren Sie den erzeugten Datensatz mithilfe eines Barplots, eines Barplots und eines Lineplots.
17. Bestimmen Sie Betaparameterschätzer für den von Ihnen erzeugten Datensatz im Modell der additiven 2×2 ZVA mit Referenzgruppe und im Modell der 2×2 mit Interaktion und Referenzgruppe. Erläutern Sie die Gemeinsamkeiten und Unterschiede der Betaparameterschätzer vor dem Hintergrund der von Ihnen gewählten wahren, aber unbekanntes, Parameterwerte des Datensatzes.
18. Führen Sie basierend auf dem von Ihnen erzeugten Datensatz einen F-Test für den Haupteffekt des Faktors Therapie, einen F-Test für den Haupteffekt des Faktors Alter und einen F-Test für die Interaktion der Faktoren durch. Dokumentieren und erläutern Sie Ihre Ergebnisse.



Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

Datum	Einheit	Thema
08.04.2022	Grundlagen Osterpause	(1) Regression
22.04.2022	Grundlagen	(2) Korrelation
29.04.2022	Grundlagen	(3) Matrizen
06.05.2022	Grundlagen	(4) Normalverteilungen
13.05.2022	Theorie	(5) Modellformulierung
20.05.2022	Theorie	(6) Modellschätzung
27.05.2022	Theorie	(7) Modellevaluation
03.06.2021	Anwendung	(8) Studiendesign
10.06.2021	Anwendung	(9) T-Tests
17.06.2021	Anwendung	(10) Einfaktorielle Varianzanalyse
24.06.2022	Anwendung	(11) Zweifaktorielle Varianzanalyse
01.07.2022	Anwendung	(12) Multiple Regression
11.07.2022	Q & A	Online 14 - 17 Uhr
14.07.2022	Klausur	G16-H5 11 - 12 Uhr
März 2023	Klausurwiederholungstermin	

(12) Multiple Regression

Faktorielle und Parametrische ALM Designs

Faktorielle ALM Designs

- Designmatrizen mit 1en und 0en, manchmal -1 en.
- Betaparameter repräsentieren Gruppenerwartungswerte.
- Betaparameterschätzer repräsentieren Gruppenstichprobenmittel.
- \Rightarrow T-Tests, Einfaktorielle Varianzanalyse, Mehrfaktorielle Varianzanalyse

Parametrische ALM Designs

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Regressor-Daten Kovarianzen.
- Es besteht ein enger Bezug zur Theorie der Korrelation.
- \Rightarrow Einfache lineare Regression, Multiple lineare Regression

Faktoriell-parametrische ALM Designs

- Designmatrizen mit mehreren faktoriellen und parametrischen Werten.
- Die parametrischen Regressoren werden oft als kontrollierte Kovariaten betrachtet.
- \Rightarrow Kovarianzanalyse

ALM Designs als Hypothesentestverfahren*

Testen von Unterschiedshypothesen

- T-Tests
- Einfaktorielle Varianzanalyse
- Mehrfaktorielle Varianzanalyse
- Kovarianzanalyse

Testen von Zusammenhangshypothesen

- Einfache lineare Regression/Korrelation
- Multiple lineare Regression/Multiple Korrelation

*Diese Sichtweise durch den Lehrenden nicht favorisiert.

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Anwendungsszenario

- Generalisierung der einfachen linearen Regression zu mehr als einer unabhängigen Variable.
- Eine univariate abhängige Variable bestimmt an randomisierten experimentellen Einheiten.
- Zwei oder mehr "kontinuierliche" unabhängige Variablen.
- Die unabhängigen Variablen heißen Regressoren, Prädiktoren, Kovariaten oder Features.

Ziele

- Quantifizierung des Erklärungspotentials der Variation der AV durch die Variation der UVs.
- Quantifizierung des Einflusses einzelner UVs auf die AV im Kontext anderer UVs.
- Prädiktion von AV Werten aus UVs Werten nach Parameterschätzung.

Anwendungsbeispiel

- BDI Differenzwerte in Abhängigkeit von Therapiedauer und Alter

Anwendungsszenario

Beispieldatensatz

$n = 100$

ID	Age	Therapy	BDI
1	50	16	9
2	38	13	9
3	46	16	10
4	62	17	3
5	25	23	38
6	34	23	29
7	36	24	31
8	36	18	21
9	57	20	20
10	46	17	16
11	59	21	18
12	54	22	24
13	27	14	20
14	56	18	12
15	41	20	31
16	46	23	30
17	23	15	23
18	36	22	27
19	44	19	23
20	70	20	11
21	72	13	-3
22	57	16	12
23	67	16	3
24	41	19	23
25	44	14	15

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Definition (Modell der multiplen Regression)

y_i mit $i = 1, \dots, n$ sei die Zufallsvariable, die den i ten Wert einer abhängigen Variable modelliert. Dann hat das *Modell der multiplen Regression* die strukturelle Form

$$y_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ und } \sigma^2 > 0, \quad (1)$$

wobei $x_{ij} \in \mathbb{R}$ mit $1 \leq i \leq n$ und $1 \leq j \leq p$ den i ten Wert der j ten unabhängigen Variable bezeichnet. Die unabhängigen Variablen werden auch *Regressoren*, *Prädiktoren*, *Kovariaten* oder *Features* genannt. Mit

$$x_i := (x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p \text{ und } \beta := (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p \quad (2)$$

hat das Modell der multiplen Regression die Datenverteilungsform

$$y_i \sim N(\mu_i, \sigma^2) \text{ u.v. für } i = 1, \dots, n, \text{ wobei } \mu_i := x_i^T \beta. \quad (3)$$

In diesem Zusammenhang wird $x_i \in \mathbb{R}^p$ auch als *iter Featurevektor* bezeichnet. Die Designmatrixform des Modells der multiplen Regression schließlich ist gegeben durch

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (4)$$

mit

$$y := (y_1, \dots, y_n)^T, X := (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p} \in \mathbb{R}^{n \times p}, \beta := (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p \text{ und } \sigma^2 > 0. \quad (5)$$

Bemerkung

- Das Modell der multiplen Regression und die allgemeine Form des ALMs sind identisch.

Beispieldatensatzerzeugung

```
# Datensimulation
library(MASS)
set.seed(10)
n      = 100
p      = 3
x_1    = round(runif(n,20,80))
x_2    = round(runif(n,12,24))
X      = matrix(c(rep(1,n),x_1,x_2), nrow = n)
I_n    = diag(n)
beta   = matrix(c(5,-.5,2), nrow = p)
sigsqr = 10
y      = mvrnorm(1, X %*% beta, sigsqr*I_n)

# Multivariate Normalverteilung
# reproduzierbare Daten
# Anzahl Datenpunkte
# Anzahl Parameter
# Regressorwerte Alter
# Regressorwerte Therapiedauer
# Designmatrix
# Identitätsmatrix
# Betaparametervektor
# Varianzparameter
# eine Realisierung eines n-dimensionalen ZVs

# Dataframeformatierung
library(writexl)
D      = data.frame("ID" = 1:n)
D$Age  = x_1
D$Duration = x_2
D$BDI  = y

# Excel Output
# Dataframe Initialisierung und ID Variable
# Alter
# Therapiedauer
# PrePost-BDI Differenzwerte

# Datenspeicherung
write_xlsx(D, file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.xlsx"))
write_csv(D, file = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv"))
```

Beispieldatenvisualisierung

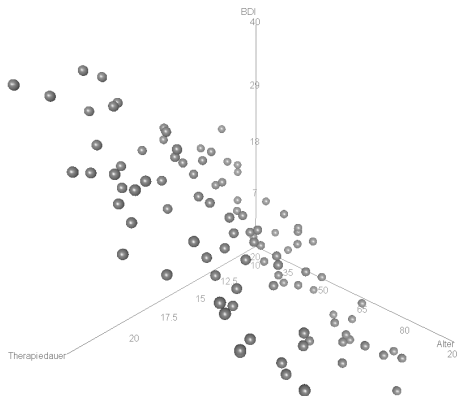
```
# Dateneinlesen
fname      = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv")
D          = read.table(fname, sep = ",", header = TRUE)

# Open GL Visualisierung mit car package, siehe ?scatter3d für Details
library(car)
scatter3d(
  D$Age,
  D$BDI,
  D$Duration,
  xlab      = "Alter",
  ylab      = "BDI",
  zlab      = "Therapiedauer",
  point.col = "gray40",
  axis.col  = rep("black",3),
  axis.scales = T,
  axis.ticks = T,
  surface   = F)
```

> Lade nötigen Namensraum: rgl

> Lade nötigen Namensraum: mgcv

Beispieldatenvisualisierung



Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Überblick

Der Betaparameterschätzer hat bekanntlich die Form

$$\hat{\beta} := (X^T X)^{-1} X^T y \quad (6)$$

Dabei quantifizieren in sehr grober Auflösung

- $X^T y \in \mathbb{R}^p$ die Kovariation der Regressoren mit den Daten und
- $X^T X \in \mathbb{R}^{p \times p}$ die Kovariation der Regressoren untereinander.

Damit ergibt sich für die Betaparameterschätzer also eine Interpretation als “regressorkovariations-normalisierte Regressordatenkovariation”,

$$\hat{\beta} \approx \text{Regressorkovariation}^{-1} \cdot \text{Regressordatenkovariation} \quad (7)$$

Im Folgenden wollen wir diese Intuition am Beispiel einer einfachen multiplen Regression mit einem Interzeptregressor und zwei unabhängigen Variablen Regressoren vertiefen, wobei die betreffenden Kovariationen einmal durch Stichprobenkorrelationen und einmal durch partielle Stichprobenkorrelationen quantifiziert werden sollen.

Theorem (Betaparameterschätzer und Korrelationen)

Gegeben sei ein multiples Regressionsmodell der Form

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (8)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \\ \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \\ \frac{r_{y,x_2} - r_{y,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_2}} \end{pmatrix}, \quad (9)$$

wobei für die y_i, x_{i1} und x_{i2} mit $i = 1, \dots, n$, $\bar{\cdot}$, s_{\cdot} und $r_{\cdot, \cdot}$ die entsprechenden Stichprobenmittel, Stichprobenstandardabweichungen, und Stichprobenkorrelationen bezeichnen.

Bemerkung

- In Bezug auf die Regressoren sind die Begriffe Stichprobenmittel, Stichprobenstandardabweichung, und Stichprobenkorrelation lediglich formal gemeint, nach Voraussetzung des ALMs sind die Regressorenwerte keine Realisierungen von Zufallsvariablen.

Modellschätzung

Beweis

Wir erinnern zunächst daran, dass die Form des Betaparameterschätzers bekanntlich zum System der Normalgleichungen äquivalent ist (vgl. (6) Modellschätzung),

$$\hat{\beta} = (X^T X)^{-1} X^T y \Leftrightarrow X^T X \hat{\beta} = X^T y. \quad (10)$$

Ausschreiben des Normalgleichungssystems für den hier betrachteten ALM Spezialfall ergibt dann zunächst

$$\begin{aligned} X^T X \hat{\beta} &= X^T y \\ \Leftrightarrow \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \sum_{i=1}^n y_i x_{i2} \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \sum_{i=1}^n y_i x_{i2} \end{pmatrix} \end{aligned}$$

Beweis (fortgeführt)

und damit

$$X^T X \hat{\beta} = X^T y$$
$$\Leftrightarrow \begin{pmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1}x_{i2} \\ \hat{\beta}_0 \sum_{i=1}^n x_{i2} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}x_{i2} + \hat{\beta}_2 \sum_{i=1}^n x_{i2}^2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \sum_{i=1}^n y_i x_{i2} \end{pmatrix}$$

Aus der Gleichung der ersten Vektorkomponenten folgt dann direkt die Form von $\hat{\beta}_0$ mit

$$\begin{aligned} \sum_{i=1}^n y_i &= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} \\ \Leftrightarrow \frac{1}{n} \sum_{i=1}^n y_i &= \hat{\beta}_0 + \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \frac{1}{n} \sum_{i=1}^n x_{i2} \\ \Leftrightarrow \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \end{aligned} \tag{11}$$

Beweis (fortgeführt)

Einsetzen dieser Form von $\hat{\beta}_0$ in die Gleichung der zweiten Vektorkomponenten ergibt dann

$$\hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n y_i x_{i1}$$

$$(\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2) \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n y_i x_{i1}$$

$$\bar{y} \sum_{i=1}^n x_{i1} - \hat{\beta}_1 \bar{x}_1 \sum_{i=1}^n x_{i1} - \hat{\beta}_2 \bar{x}_2 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n y_i x_{i1}$$

$$\hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 - \hat{\beta}_1 \bar{x}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} - \hat{\beta}_2 \bar{x}_2 \sum_{i=1}^n x_{i1} = \sum_{i=1}^n y_i x_{i1} - \bar{y} \sum_{i=1}^n x_{i1}$$

$$\hat{\beta}_1 \left(\sum_{i=1}^n x_{i1}^2 - \bar{x}_1 \sum_{i=1}^n x_{i1} \right) + \hat{\beta}_2 \left(\sum_{i=1}^n x_{i1} x_{i2} - \bar{x}_2 \sum_{i=1}^n x_{i1} \right) = \sum_{i=1}^n y_i x_{i1} - \bar{y} \sum_{i=1}^n x_{i1}$$

Beweis (fortgeführt)

Im Beweis des Theorems zur Ausgleichsgerade (vgl. (1) Regression) haben wir gesehen, dass

$$\begin{aligned} \sum_{i=1}^n x_{i1}x_{i1} - \bar{x}_1 \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1) \\ \sum_{i=1}^n x_{i1}x_{i2} - \bar{x}_2 \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) \\ \sum_{i=1}^n y_i x_{i1} - \bar{y} \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1) \end{aligned} \tag{12}$$

Beweis (fortgeführt)

Es ergibt sich also, dass

$$\begin{aligned} \hat{\beta}_1 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1) + \hat{\beta}_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) &= \sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1) \\ \hat{\beta}_1 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1)}{n-1} + \hat{\beta}_2 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{n-1} &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1)}{n-1} \end{aligned} \quad (13)$$

Mit den Definitionen von Stichprobenstandardabweichung und -korrelation folgt dann weiter

$$\begin{aligned} \hat{\beta}_1 s_{x_1} s_{x_1} + \hat{\beta}_2 c_{x_1, x_2} &= c_{y, x_1} \\ \hat{\beta}_1 \frac{s_{x_1} s_{x_1}}{s_y s_{x_1}} + \hat{\beta}_2 \frac{c_{x_1, x_2}}{s_y s_{x_1}} &= \frac{c_{y, x_1}}{s_y s_{x_1}} \\ \hat{\beta}_1 \frac{s_{x_1}}{s_y} + \hat{\beta}_2 \frac{c_{x_1, x_2}}{s_y s_{x_1}} &= r_{y, x_1} \\ \hat{\beta}_1 \frac{s_{x_1}}{s_y} + \hat{\beta}_2 \frac{c_{x_1, x_2} s_{x_2}}{s_y s_{x_1} s_{x_2}} &= r_{y, x_1} \\ \hat{\beta}_1 \frac{s_{x_1}}{s_y} + \hat{\beta}_2 \frac{s_{x_2}}{s_y} r_{x_1, x_2} &= r_{y, x_1} \end{aligned} \quad (14)$$

Beweis (fortgeführt)

Definition von

$$b_j := \frac{s_{x_j}}{s_y}, j = 1, 2 \quad (15)$$

erlaubt dann die Schreibweise

$$b_1 + b_2 r_{x_1, x_2} = r_{y, x_1} \quad (16)$$

Schließlich folgt analog durch Vertauschen der Subskripte aus der Gleichung der dritten Vektorkomponenten

$$b_1 r_{x_1, x_2} + b_2 = r_{y, x_2} \quad (17)$$

Insgesamt haben wir also gesehen, dass die Definition des Betaparameterschätzers im vorliegenden ALM Spezialfall ergibt, dass mit

$$\hat{\beta}_j = b_j \frac{s_y}{s_{x_j}}, j = 1, 2 \quad (18)$$

gilt, dass

$$\begin{aligned} r_{y, x_1} &= b_1 + b_2 r_{x_1, x_2} \\ r_{y, x_2} &= b_1 r_{x_1, x_2} + b_2 \end{aligned} \quad (19)$$

Beweis (fortgeführt)

Damit folgt aus der zweiten Gleichung dann sofort

$$b_2 = r_{y,x_2} - b_1 r_{x_1,x_2}. \quad (20)$$

Einsetzen in die erste Gleichung ergibt dann

$$\begin{aligned} b_1 + (r_{y,x_2} - b_1 r_{x_1,x_2}) r_{x_1,x_2} &= r_{y,x_1} \\ \Leftrightarrow b_1 + r_{y,x_2} r_{x_1,x_2} - b_1 r_{x_1,x_2}^2 &= r_{y,x_1} \\ \Leftrightarrow r_{y,x_2} r_{x_1,x_2} + b_1 (1 - r_{x_1,x_2}^2) &= r_{y,x_1} \\ \Leftrightarrow b_1 (1 - r_{x_1,x_2}^2) &= r_{y,x_1} - r_{y,x_2} r_{x_1,x_2} \\ \Leftrightarrow b_1 &= \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \end{aligned} \quad (21)$$

Modellschätzung

Beweis (fortgeführt)

Für b_2 ergibt sich damit weiterhin

$$\begin{aligned} b_2 &= r_{y,x_2} - b_1 r_{x_1,x_2} \\ \Leftrightarrow b_2 &= r_{y,x_2} - \left(\frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \right) r_{x_1,x_2} \\ \Leftrightarrow b_2 &= \frac{r_{y,x_2} (1 - r_{x_1,x_2}^2)}{1 - r_{x_1,x_2}^2} - \frac{r_{y,x_1} r_{x_1,x_2} - r_{y,x_2} r_{x_1,x_2}^2}{1 - r_{x_1,x_2}^2} \\ \Leftrightarrow b_2 &= \frac{r_{y,x_2} - r_{y,x_2} r_{x_1,x_2}^2 - r_{y,x_1} r_{x_1,x_2} + r_{y,x_2} r_{x_1,x_2}^2}{1 - r_{x_1,x_2}^2} \\ \Leftrightarrow b_2 &= \frac{r_{y,x_2} - r_{y,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \end{aligned} \tag{22}$$

Damit folgen dann aber

$$\begin{aligned} \hat{\beta}_1 &= b_1 \frac{s_y}{s_{x_1}} = \left(\frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \right) \frac{s_y}{s_{x_1}} \\ \hat{\beta}_2 &= b_2 \frac{s_y}{s_{x_2}} = \left(\frac{r_{y,x_2} - r_{y,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \right) \frac{s_y}{s_{x_2}} \end{aligned} \tag{23}$$

und es ist alles gezeigt. □

Modellschätzung

Anwendungsbeispiel

```
# Dateneinlesen
fname = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv")
D      = read.table(fname, sep = ",", header = TRUE)      # Datensatz

# Modellschätzung
y      = D$BDI                                           # Abhängige Variable
n      = length(y)                                       # Anzahl Datenpunkte
X      = matrix(c(rep(1,n), D$Age, D$Duration), nrow = n) # Designmatrix
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y             # Betaparameterschätzer
eps_hat  = y - X %*% beta_hat                           # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)           # Varianzparameterschätzer

# Betaparameterschätzer aus Stichprobenmittel, -standardabweichungen und -korrelationen
y12    = cbind(y,X[,-1])                                 # y, x_1, x_2 Matrix
bars    = apply(y12, 2, mean)                             # Stichprobenmittel
s       = apply(y12, 2, sd)                               # Stichprobenstandardabweichungen
r       = cor(y12)                                       # Stichprobenkorrelationen
beta_hat_1 = (r[1,2] - r[1,3]*r[2,3]) / (1 - r[2,3]^2) * (s[1]/s[2]) # \hat{\beta}_1
beta_hat_2 = (r[1,3] - r[1,2]*r[2,3]) / (1 - r[2,3]^2) * (s[1]/s[3]) # \hat{\beta}_2
beta_hat_0 = bars[1] - beta_hat_1*bars[2] - beta_hat_2*bars[3] # \hat{\beta}_0

# Ausgabe
cat("beta_hat ALM-Schätzer      : " , beta_hat,
    "\nbeta_hat Deskriptivstatistiken :", c(beta_hat_0,beta_hat_1,beta_hat_2))
```

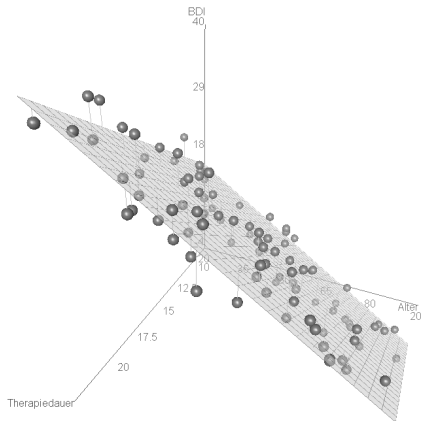
```
> beta_hat ALM-Schätzer      : 5.42 -0.481 1.91
> beta_hat Deskriptivstatistiken : 5.42 -0.481 1.91
```

Beispieldatenvisualisierung

```
# Dateneinlesen
fname      = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv")
D          = read.table(fname, sep = ",", header = TRUE)      # Datensatz

# Open GL Visualisierung mit car package, siehe ?scatter3d für Details
library(car)
scatter3d(
  D$Age,
  D$BDI,
  D$Duration,
  xlab      = "Alter",
  ylab      = "BDI",
  zlab      = "Therapiedauer",
  point.col = "gray40",
  axis.col  = rep("black",3),
  axis.scales = T,
  axis.ticks = T,
  surface   = T,
  surface.col = "gray70",
  neg.res.col = "gray70",
  pos.res.col = "gray70")
```


Beispieldatenvisualisierung



Theorem (Betaparameterschätzer und partielle Korrelationen)

Gegeben sei ein multiples Regressionsmodell der Form

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (24)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \\ r_{y, x_1 | x_2} \sqrt{\frac{1 - r_{y, x_2}^2}{1 - r_{x_1, x_2}^2}} \frac{s_y}{s_{x_1}} \\ r_{y, x_2 | x_1} \sqrt{\frac{1 - r_{y, x_1}^2}{1 - r_{x_2, x_1}^2}} \frac{s_y}{s_{x_2}} \end{pmatrix}, \quad (25)$$

wobei für $1 \leq k, l \leq 2$ und $i = 1, \dots, n$

- $r_{y, x_k | x_l}$ die partielle Stichprobenkorrelation der y_i und x_{ik} gegeben die x_{il} ist,
- r_{y, x_k} die Stichprobenkorrelation der y_i und x_{ik} ist, und
- r_{x_k, x_l} die Stichprobenkorrelation der x_{ik} und x_{il} ist.

Bemerkungen

- Im Allgemeinen gilt für $1 \leq i, l \leq k$, dass $\hat{\beta}_k \neq r_{y, x_k | x_l}$.
- Betaparameterschätzer sind also im Allgemeinen keine partiellen Stichprobenkorrelationen.
- $\hat{\beta}_k = r_{y, x_k | x_l}$ für $1 \leq i, l \leq k$ gilt genau dann, wenn $s_y = s_{x_1} = s_{x_2}$ und zudem
 - $r_{y, x_k} = r_{x_k, x_l} = 0$, wenn also die Stichprobenkorrelationen der Daten und der Werte des zweiten Regressors, sowie die Stichprobenkorrelation der Werte der beiden Regressoren gleich Null sind. Dies kann der Fall sein, wenn einer der Regressoren die Daten "sehr gut erklärt" und der andere Regressor von dem ersten "sehr verschieden" ist.
 - $|r_{y, x_l}| = |r_{x_k, x_l}|$, wenn also die obige Stichprobenkorrelationen dem Betrage nach gleich sind. Dies ist vermutlich selten der Fall.

Modellschätzung

Beweis

Wir betrachten $\hat{\beta}_1$, das Resultat für $\hat{\beta}_2$ folgt dann durch Vertauschen der Indizes. Wir haben in vorherigem Theorem gesehen, dass

$$\hat{\beta}_1 = \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \quad (26)$$

Weiterhin haben wir in (2) Korrelation gesehen, dass unter der Annahme der multivariaten Normalverteilung von y, x_1, x_2 ein Schätzer für die partielle Korrelation von y und x_1 gegeben x_2 durch

$$r_{y,x_1|x_2} = \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \quad (27)$$

gegeben ist. Für $\hat{\beta}_1$ ergibt sich somit

$$\begin{aligned} \hat{\beta}_1 &= \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \\ \Leftrightarrow (1 - r_{x_1,x_2}^2) \hat{\beta}_1 &= (r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}) \frac{s_y}{s_{x_1}} \\ \Leftrightarrow \frac{1 - r_{x_1,x_2}^2}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \hat{\beta}_1 &= \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}} \quad (28) \\ \Leftrightarrow \frac{1 - r_{x_1,x_2}^2}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \hat{\beta}_1 &= r_{y,x_1|x_2} \frac{s_y}{s_{x_1}} \end{aligned}$$

Beweis

und damit weiter

$$\begin{aligned}\hat{\beta}_1 &= r_{y,x_1|x_2} \frac{\sqrt{1-r_{y,x_2}^2} \sqrt{1-r_{x_1,x_2}^2}}{1-r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \\ \Leftrightarrow \hat{\beta}_1 &= r_{y,x_1|x_2} \frac{\sqrt{1-r_{y,x_2}^2} \sqrt{1-r_{x_1,x_2}^2}}{\left(\sqrt{1-r_{x_1,x_2}^2}\right)^2} \frac{s_y}{s_{x_1}} \\ \Leftrightarrow \hat{\beta}_1 &= r_{y,x_1|x_2} \frac{\sqrt{1-r_{y,x_2}^2}}{\sqrt{1-r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}} \\ \Leftrightarrow \hat{\beta}_1 &= r_{y,x_1|x_2} \sqrt{\frac{1-r_{y,x_2}^2}{1-r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}}\end{aligned}\tag{29}$$

Modellschätzung

Anwendungsbeispiel

```
# Dateneinlesen
fname = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE) # Datensatz

# Modellschätzung
y = D$BDI # Abhängige Variable
n = length(y) # Anzahl Datenpunkte
X = matrix(c(rep(1,n), D$Age, D$Duration), nrow = n) # Designmatrix
p = ncol(X) # Anzahl Parameter
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer

# Betaparameterschätzer aus partiellen Korrelationen und Korrelationen
library(ppcor) # partielle Korrelationentoolbox
y12 = cbind(y,X[,-1]) # y, x_1, x_2 Matrix
bars = apply(y12, 2, mean) # Stichprobenmittel
s = apply(y12, 2, sd) # Stichprobenstandardabweichungen
r = cor(y12) # Stichprobenkorrelationen
pr = pcor(y12) # partielle Stichprobenkorrelationen
pr = pr$estimate # partielle Stichprobenkorrelationen
beta_hat_1 = pr[1,2]*sqrt((1-r[1,3]^2)/(1-r[2,3]^2))*(s[1]/s[2]) # \hat{\beta}_1
beta_hat_2 = pr[1,3]*sqrt((1-r[1,2]^2)/(1-r[3,2]^2))*(s[1]/s[3]) # \hat{\beta}_2
beta_hat_0 = bars[1] - beta_hat_1*bars[2] - beta_hat_2*bars[3] # \hat{\beta}_0

# Ausgabe
cat("Korrelationen r(y,x_1),r(y,x_2),r(x_1,x_2) :", c(r[1,2],r[1,3],r[2,3]),
    "\nPartielle Korrelationen r(y,x_1|x_2), r(y,x_2|x_1) :", c(pr[1,2],pr[1,3]),
    "\nbeta_hat ALM Schätzer :", beta_hat,
    "\nbeta_hat aus partieller Korrelation :", c(beta_hat_0,beta_hat_1,beta_hat_2))

> Korrelationen r(y,x_1),r(y,x_2),r(x_1,x_2) : -0.726 0.644 -0.0268
> Partielle Korrelationen r(y,x_1|x_2), r(y,x_2|x_1) : -0.927 0.909
> beta_hat ALM Schätzer : 5.42 -0.481 1.91
> beta_hat aus partieller Korrelation : 5.42 -0.481 1.91
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Parameterinferenz | T-Tests

Zur Erinnerung (vgl. (7) Modellevaluation)

Theorem (T-Teststatistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (30)$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (31)$$

die Betaparameter- und Varianzparameterschätzer, respektive. Schließlich sei für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ und einen *Nullhypothesenbetaparameter* $\beta_0 \in \mathbb{R}^p$ die *T-Teststatistik* definiert als

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \quad (32)$$

Dann gilt

$$T \sim t(\delta, n - p) \text{ mit } \delta := \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} \quad (33)$$

Parameterinferenz | T-Tests

Einige mögliche Kontrastgewichtsvektoren und Nullhypothesen im Anwendungsbeispiel:

$$c = (1, 0, 0)^T \quad H_0 : \beta_1 = 0 \quad H_A : \beta_1 \neq 0$$

$$c = (0, 1, 0)^T \quad H_0 : \beta_2 = 0 \quad H_A : \beta_2 \neq 0$$

$$c = (0, 0, 1)^T \quad H_0 : \beta_3 = 0 \quad H_A : \beta_3 \neq 0$$

$$c = (0, 1, -1)^T \quad H_0 : \beta_2 - \beta_3 = 0 \quad H_A : \beta_2 - \beta_3 \neq 0$$

...

...

...

Modellevaluation

Parameterinferenz | T-Tests

```
# Dateneinlesen
fname = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE) # Datensatz

# Modellschätzung
y = D$BDI # Abhängige Variable
n = length(y) # Anzahl Datenpunkte
X = matrix(c(rep(1,n), D$Age, D$Duration), nrow = n) # Designmatrix
p = ncol(X) # Anzahl Parameter
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer

# Modellevaluation / Parameterinferenz
C = cbind(diag(p), matrix(c(0,1,-1), nrow = 3)) # Kontrastgewichtsvektoren
ste = rep(NA, ncol(C)) # Kontraststandardfehler
tee = rep(NA, ncol(C)) # T-Statistiken
pvals = rep(NA, ncol(C)) # p-Werte
for(i in 1:ncol(C)){
  c = C[,i] # Kontrastgewichtsvektor
  t_num = t(c) %*% beta_hat # Zähler der T-Statistik
  ste[i] = sqrt(sigsqr_hat * t(c) %*% solve(t(X) %*% X) %*% c) # Kontraststandardfehler/Nummer der T-Statistik
  tee[i] = t_num / ste[i] # T-Statistik
  pvals[i] = 2 * (1 - pt(abs(tee[i]), n-p)) # p-Wert
}

# Ausgabe
R = data.frame(c(beta_hat, t(C[,4] %*% beta_hat)), ste, tee, pvals)
rownames(R) = c("(Intercept)", "Age", "Therapy", "Age-Therapy")
colnames(R) = c("Estimate", "Std. Error", "t value", "Pr(>|t|)")
print(R)
```

```
> Estimate Std. Error t value Pr(>|t|)
> (Intercept) 5.422 1.9024 2.85 0.00534
> Age -0.481 0.0198 -24.33 0.00000
> Therapy 1.912 0.0893 21.41 0.00000
> Age-Therapy -2.393 0.0909 -26.32 0.00000
```

Modellinferenz | F-Tests

Zur Erinnerung (vgl. (7) Modellevaluation)

Theorem (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (34)$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \quad (35)$$

mit $p = p_1 + p_2$ gegeben. Schließlich sei

$$K := \begin{pmatrix} 0_{p_1} \\ 1_{p_2} \end{pmatrix} \in \mathbb{R}^p \quad (36)$$

ein Kontrastgewichtsvektor. Dann gilt

$$F \sim f(\delta, p_2, n - p) \text{ mit } \delta := \frac{K^T \beta \left(K^T (X^T X)^{-1} K \right)^{-1} K^T \beta}{\sigma^2} \quad (37)$$

Modellevaluation

Modellinferenz | F-Tests

$p_1 := 1$

```
# Dateneinlesen
fname = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv")
D = read.table(fname, sep = ",", header = TRUE) # Datensatz

# Modellevaluation
y = D$BDI # Abhängige Variable
n = length(y) # Anzahl Datenpunkte
X = matrix(c(rep(1,n), D$Age, D$Duration), nrow = n) # Designmatrix vollständiges Modell
p = ncol(X) # Anzahl Parameter vollständiges Modell
p_1 = 1 # Anzahl Parameter reduziertes Modell
p_2 = p - p_1 # Anzahl zusätzlicher Parameter im vollst. Modell
X_1 = X[,1:p_1] # Designmatrix reduziertes Modell
beta_hat_1 = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y # Betaparameterschätzer reduziertes Modell
beta_hat = solve(t(X) %*%X )%*%t(X) %*%y # Betaparameterschätzer vollständiges Modell
eps_hat_1 = y - X_1%*%beta_hat_1 # Residuenvektor reduziertes Modell
eps_hat = y - X%*%beta_hat # Residuenvektor vollständiges Modell
eh1_eh1 = t(eps_hat_1) %*% eps_hat_1 # RQS reduziertes Modell
eh_eh = t(eps_hat) %*% eps_hat # RQS vollständiges Modell
sigsqr_hat = eh_eh/(n-p) # Varianzparameterschätzer vollst. Modell
f = ((eh1_eh1-eh_eh)/p_2)/sigsqr_hat # F-Statistik
pval = 1 - pf(f,p_2,n-p) # p-Wert

# Ausgabe
cat("F-statistic:", f, "on", p_2, "and", n-p, "DF", "p-value: ", paste(pval))
```

```
> F-statistic: 540 on 2 and 97 DF p-value: 0
```

Modellformulierung, Modellschätzung und Modellevaluation mit R

```
fname = file.path(getwd(), "12_Daten", "12_Multiple_Regression_Daten.csv") # Datensatzdatei
D      = read.table(fname, sep = ",", header = TRUE)                       # Datensatzeinlesen
alm    = lm(BDI ~ Age + Therapy, data = D)                                # Modellformulierung und Modellschätzung
summary(alm)
```

```
>
> Call:
> lm(formula = BDI ~ Age + Therapy, data = D)
>
> Residuals:
>   Min     1Q   Median     3Q      Max
> -7.178 -2.165  0.438  2.585  7.119
>
> Coefficients:
>             Estimate Std. Error t value Pr(>|t|)
> (Intercept)  5.4225     1.9024   2.85  0.0053 **
> Age         -0.4815     0.0198  -24.33 <2e-16 ***
> Therapy      1.9119     0.0893   21.41 <2e-16 ***
> ---
> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> Residual standard error: 3.07 on 97 degrees of freedom
> Multiple R-squared:  0.918,    Adjusted R-squared:  0.916
> F-statistic: 540 on 2 and 97 DF,  p-value: <2e-16
```

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Allgemeines Lineares Modell SoSe 2023

- Konfidenzintervalle
- Allgemeine Kontrasttheorie
- Kovarianzanalyse

Weiterführende Theorie des Allgemeinen Linearen Modells

Relaxation der Unabhängigkeitsannahme der Fehlerterme

⇒ Generalized Least Squares, Repeated-Measures Designs, ...

Modellierung von Beta- und Varianzparametern als Zufallsvariablen

⇒ Hierarchische lineare Modelle, linear mixed models, Bayesian estimation, Varianzkomponentenschätzung, ...

Nichtlineare Transformationen von Erwartungswertparametern

⇒ Generalisierte lineare Modelle, logistische Regression, neuronale Netze, ...

Multivariate Erweiterung der Datenvariable

⇒ Multivariate ALMs, Faktoranalyse, Strukturgleichungsmodelle, ...

Zeitliche Erweiterung der Datenvariable

⇒ Linear Gaussian State Space Models, Kalman Filter, Bayesian Filtering, ...

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Ausblick

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario und die Ziele der multiplen Regression.
2. Definieren Sie das Modell der multiplen Regression.
3. Erläutern Sie die Begriffe Regressor, Prädiktor, Kovariate und Feature im Rahmen der multiplen Regression.
4. Erläutern Sie, warum $\hat{\beta} \approx \text{Regressorkovarianz}^{-1} \text{Regressordatenkovarianz}$ gilt.
5. Erläutern Sie den Zusammenhang zwischen Betaparameterschätzern und partieller Korrelation in einem multiplen Regressionmodell mit Interzeptprädiktor und zwei kontinuierlichen Prädiktoren anhand der Formel

$$\hat{\beta}_1 = r_{y, x_1 | x_2} \sqrt{\frac{1 - r_{y, x_2}^2}{1 - r_{x_1, x_2}^2}} \frac{s_y}{s_{x_1}}. \quad (38)$$

6. $X \in \mathbb{R}^{n \times 2}$ sei die Designmatrix eines multiplen Regressionsmodells mit zwei Prädiktoren und Betaparametervektor $\beta := (\beta_1, \beta_2)^T$. Geben Sie den Kontrastgewichtsvektor an, um die Nullhypothese $H_0 : \beta_1 = \beta_2$ mithilfe der T-Statistik zu testen.
7. Simulieren Sie einen Datensatz eines multiplen Regressionsmodells mit Interzept und zwei kontinuierlichen Regressoren $x_1, x_2 \in \mathbb{R}^n$, wobei $x_{i2} := ax_{i1} + \xi_i$ mit $\xi_i \sim N(0, \sigma_\xi^2)$ für $i = 1, \dots, n$ sein soll. Wählen Sie für die Simulation des Datensatzes $y \in \mathbb{R}^n$ den wahren, aber unbekanntem, Betaparametervektor $\beta = (0, 1, 0)^T$ und testen Sie die Nullhypothesen $H_0 : \beta_j = 0$ für $j = 0, 1, 2$. Erläutern Sie Ihre Ergebnisse. Wiederholen Sie Analyse für den wahren, aber unbekanntem, Betaparametervektor $\beta = (0, 0, 1)^T$.