

# Allgemeines Lineares Modell

Dirk Ostwald

Joram Soch

2024-07-03

# Inhaltsverzeichnis

<b>Vorwort</b>	<b>3</b>
<b>1 Regression</b>	<b>4</b>
1.1 Methode der kleinsten Quadrate . . . . .	5
1.2 Einfache lineare Regression . . . . .	12
1.3 Literaturhinweise . . . . .	16
1.4 Selbstkontrollfragen . . . . .	16
<b>2 Korrelation</b>	<b>18</b>
2.1 Grundlagen . . . . .	18
2.2 Korrelation und Bestimmtheitsmaß . . . . .	20
2.3 Korrelation und linear-affine Abhängigkeit . . . . .	26
2.4 Literaturhinweise . . . . .	29
2.5 Selbstkontrollfragen . . . . .	29
<b>3 Matrizen</b>	<b>30</b>
3.1 Definition . . . . .	30
3.2 Operationen . . . . .	32
3.3 Matrixmultiplikation . . . . .	38
3.4 Matrixinversion . . . . .	43
3.5 Determinanten . . . . .	47
3.6 Spezielle Matrizen . . . . .	51
3.7 Literaturhinweise . . . . .	55
3.8 Selbstkontrollfragen . . . . .	55
<b>4 Normalverteilungen</b>	<b>57</b>
4.1 Konstruktion . . . . .	57
4.2 Definition . . . . .	60
4.3 Transformationen . . . . .	62
4.4 Sphärische Verteilungen . . . . .	63
4.5 Marginale und bedingte Verteilungen . . . . .	64
4.6 Literaturhinweise . . . . .	67
4.7 Selbstkontrollfragen . . . . .	68
<b>5 Modellformulierung</b>	<b>69</b>
5.1 Allgemeine Theorie . . . . .	69
5.2 Identifizierbarkeit und Schätzbarkeit . . . . .	73
5.3 Designspektrum . . . . .	75
5.4 Literaturhinweise . . . . .	76
5.5 Selbstkontrollfragen . . . . .	76
<b>6 Parameterschätzung</b>	<b>77</b>
6.1 Betaparameterschätzung . . . . .	77

6.2	Varianzparameterschätzung . . . . .	80
6.3	Unabhängig identisch normalverteilte Zufallsvariablen . . . . .	81
6.4	Einfache lineare Regression . . . . .	82
6.5	Frequentistische Schätzerverteilungen . . . . .	86
6.6	Literaturhinweise . . . . .	90
6.7	Selbstkontrollfragen . . . . .	90
<b>7</b>	<b>T-Statistiken</b>	<b>91</b>
7.1	Definition und Beispiele . . . . .	91
7.2	Konfidenzintervalle für Betaparameterkomponenten . . . . .	96
7.3	Literaturhinweise . . . . .	99
7.4	Selbstkontrollfragen . . . . .	99
<b>8</b>	<b>F-Statistiken</b>	<b>101</b>
8.1	Likelihood-Quotienten-Statistiken . . . . .	101
8.2	Definition und Verteilung . . . . .	103
8.3	Literaturhinweise . . . . .	108
8.4	Selbstkontrollfragen . . . . .	109
<b>9</b>	<b>T-Tests</b>	<b>110</b>
9.1	Einstichproben-T-Tests . . . . .	110
9.2	Zweistichproben-T-Tests . . . . .	114
9.3	Literaturhinweise . . . . .	119
9.4	Selbstkontrollfragen . . . . .	119
<b>10</b>	<b>Einfaktorielle Varianzanalyse</b>	<b>121</b>
10.1	Anwendungsszenario . . . . .	121
10.2	Anwendungsbeispiel . . . . .	121
10.3	Modellformulierung . . . . .	122
10.4	Modellschätzung . . . . .	127
10.5	Modellevaluation . . . . .	130
10.6	Literaturhinweise . . . . .	140
10.7	Selbstkontrollfragen . . . . .	140
<b>11</b>	<b>Zweifaktorielle Varianzanalyse</b>	<b>142</b>
11.1	Anwendungsszenario . . . . .	142
11.2	Anwendungsbeispiel . . . . .	144
11.3	Modellformulierung . . . . .	147
11.4	Modellschätzung . . . . .	155
11.5	Modellevaluation . . . . .	160
11.6	Literaturhinweise . . . . .	165
11.7	Selbstkontrollfragen . . . . .	166
<b>12</b>	<b>Partielle Korrelation</b>	<b>167</b>
12.1	Motivation . . . . .	167
12.2	Bedingte Korelation . . . . .	169
12.3	Partielle Korrelation . . . . .	172
12.4	Literaturhinweise . . . . .	177
12.5	Selbstkontrollfragen . . . . .	177

---

<b>13 Multiple Regression</b>	<b>178</b>
13.1 Anwendungsszenario . . . . .	178
13.2 Modellformulierung . . . . .	179
13.3 Modellschätzung . . . . .	180
13.4 Modellevaluation . . . . .	187
13.5 Literaturhinweise . . . . .	189
13.6 Selbstkontrollfragen . . . . .	189
<b>14 Kovarianzanalyse</b>	<b>191</b>
14.1 Vorbemerkungen . . . . .	191
14.2 Additive Kovarianzanalyse . . . . .	192
14.3 Kovarianzanalyse mit Interaktion . . . . .	196
14.4 Selbstkontrollfragen . . . . .	201
<b>Referenzen</b>	<b>202</b>

# Vorwort

Die vorliegenden Aufzeichnungen bilden das Skript zur Vorlesung *B1. Inferenzstatistik – Allgemeines Lineares Modell* am [Lehrstuhl für Methodenlehre I](#) des [Instituts für Psychologie \(IPSY\)](#) der [Otto-von-Guericke Universität Magdeburg \(OvGU\)](#). Es basiert auf dem Skript, das von Dirk Ostwald für die [Vorlesung im Sommersemester 2023](#) geschrieben wurde, und berücksichtigt Änderungen, die von Joram Soch für die [Vorlesung im Sommersemester 2024](#) vorgenommen wurden.

Das Allgemeine Lineare Modell (ALM) ist ein integrativer Zugang zu einer Reihe typischer inferenzstatistischer Verfahren wie einfacher und multipler linearer Regression, Einstichproben- und Zweistichproben-T-Tests, einfaktorieller und zweifaktorieller Varianzanalysen sowie Kovarianzanalysen. Der Kurs gibt eine Einführung in das ALM und behandelt dieses im Sinne der [Wahrscheinlichkeitstheorie und Frequentistischen Inferenz](#). Das ALM bildet weiterhin eine zentrale Grundlage für fortgeschrittene Methoden der probabilistischen Modellierung sowie zu Verfahren der wahrscheinlichkeitsfreien Modellierung im Rahmen maschineller Lernverfahren.

# 1 Regression

Fundamentales Ziel von Regressionsanalysen ist es, Beziehungen zwischen unabhängigen und abhängigen Variablen zu modellieren. Ein zentrales Thema dabei ist die Anpassung von Funktionen an beobachtete Datensätze. Mit dem Begriff der Ausgleichsgerade im Rahmen der Methode der kleinsten Quadrate und dem Begriff der einfachen linearen Regression wollen wir uns in diesem Abschnitt diesen zentralen Themen der probabilistischen Datenmodellierung schrittweise nähern. Dabei unterscheiden sich die Konzepte von Ausgleichsgerade und einfacher linearer Regression in einem zentralen Aspekt: Bei der Ausgleichsgerade werden unabhängige und abhängige Variable nicht als Zufallsvariablen modelliert, im Rahmen der einfachen linearen Regression nimmt die abhängige Variable dann die Form einer Zufallsvariablen an. Im Kontext der Korrelation schließlich werden sowohl abhängige als auch unabhängige Variable als Zufallsvariablen modelliert.

Um die Konzepte dieses Abschnittes zu verdeutlichen, betrachten wir einen Beispieldatensatz, in dem die Anzahl an Psychotherapiestunden als unabhängige Variable  $x$  der Symptomreduktion einer Gruppe von  $n = 20$  Patient:innen als abhängige Variable  $y$  gegenüber gestellt wird (Abbildung 1.1). Die visuelle Inspektion dieses Datensatzes legt nahe, dass ein Mehr an Therapiestunden ein Mehr an Symptomreduktion impliziert. Ziel der Methode der kleinsten Quadrate und der einfachen linearen Regression ist es, diesen intuitiven funktionalen Zusammenhang zwischen unabhängiger und abhängiger Variable auf eine quantitative Basis zu stellen.

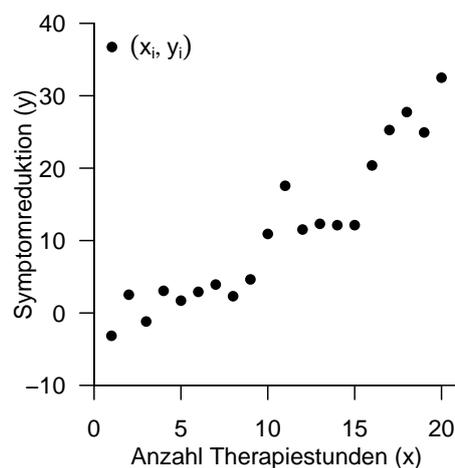


Abbildung 1.1 Beispieldatensatz.

## 1.1 Methode der kleinsten Quadrate

Wir definieren zunächst den Begriff der *Gerade*

**Definition 1.1** (Gerade). Sei  $\beta := (\beta_0, \beta_1)^T \in \mathbb{R}^2$  ein zweidimensionaler Vektor reeller Zahlen. Dann ist eine Gerade durch folgende linear-affine Funktion gegeben:

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \beta_0 + \beta_1 x . \quad (1.1)$$

•

Dies erlaubt es uns, den Begriff der *Ausgleichsgerade* einzuführen.

**Definition 1.2** (Ausgleichsgerade).  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  sei ein Datensatz. Weiterhin sei  $q(\beta)$  die Summe der quadrierten vertikalen Abweichungen der  $y_i$  von den Funktionswerten  $f_\beta(x_i)$ :

$$q : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^n (y_i - f_\beta(x_i))^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 . \quad (1.2)$$

Dann heißt die Gerade  $f_\beta$ , für die die Funktion  $q$  ihr Minimum annimmt, die *Ausgleichsgerade für den Datensatz*  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ .

•

Bei der Ausgleichsgerade handelt es sich also um eine linear-affine Funktion der Form

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \beta_0 + \beta_1 x . \quad (1.3)$$

Abbildung 1.2 zeigt drei durch jeweils andere Werte von  $\beta_0$  und  $\beta_1$  parameterisierte linear-affine Funktionen zusammen mit der Wertemenge des Beispieldatensatzes.

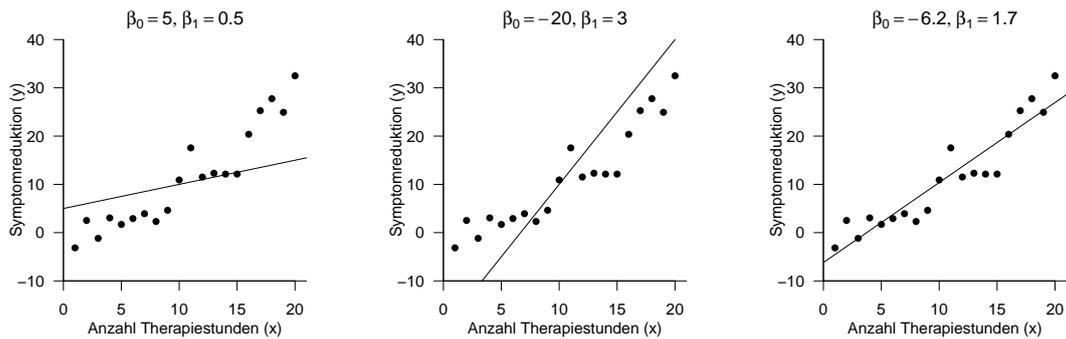
Wie bei allen linear-affinen Funktionen entspricht bei  $f_\beta$  der Wert von  $\beta_0$  dem Wert, den  $f_\beta$  für  $x = 0$  annimmt,

$$f_\beta(0) = \beta_0 + \beta_1 \cdot 0 = \beta_0 , \quad (1.4)$$

und damit graphisch dem Schnittpunkt des Funktionsgraphen mit der  $y$ -Achse. Da  $\beta_0$  damit dem Versatz (engl. *offset*) des Funktionsgraphen von  $y = 0$  an der Stelle  $x = 0$  entspricht, nennt man  $\beta_0$  auch häufig den *Offsetparameter*. Analog entspricht wie bei allen linear-affinen Funktionen der Wert von  $\beta_1$  dem Wert der Funktionswertdifferenz pro Argumenteinheitsdifferenz. Beispielsweise gilt etwa für  $\beta_0 = 5$  und  $\beta_1 = 0.5$ , dass

$$\begin{aligned} f_\beta(2) - f_\beta(1) &= (5 + 0.5 \cdot 2) - (5 + 0.5 \cdot 1) = 1 - 0.5 = 0.5 \\ f_\beta(9) - f_\beta(8) &= (5 + 0.5 \cdot 9) - (5 + 0.5 \cdot 8) = 9.5 - 8 = 0.5 \end{aligned} \quad (1.5)$$

Für eine Argumentdifferenz von 1 ergibt sich also eine Funktionswertdifferenz von 0.5.  $\beta_1$  repräsentiert also die Stärke der Änderung der Funktionswerte pro Argumenteinheitsdifferenz und damit die Steigung (engl. *slope*) des Graphen der linear-affinen Funktion. Entsprechend wird  $\beta_1$  *Steigungsparameter* oder *Slopeparameter* genannt.



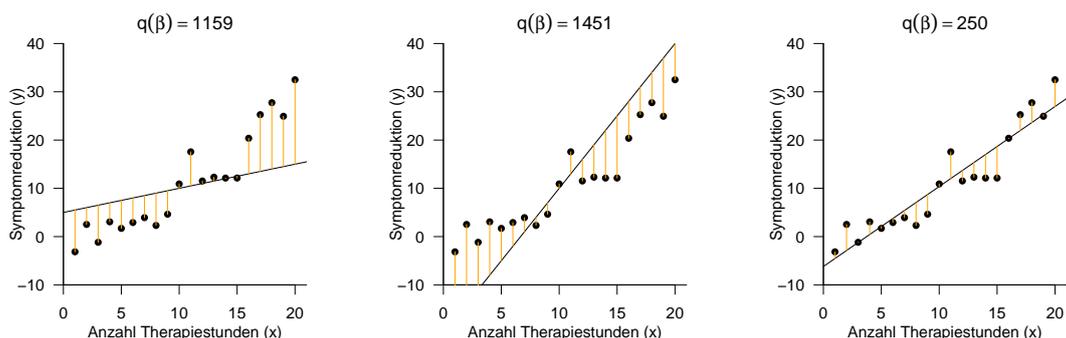
**Abbildung 1.2** Linear-affine Funktionen mit unterschiedlichen Parameterwerten vor dem Hintergrund des Beispieldatensatzes.

Nach Definition ist die Ausgleichsgerade nun allerdings nicht eine beliebige linear-affine Funktion der Form  $f_\beta$ , sondern eben jene, die für einen gegebenen Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  die Summe der quadrierten vertikalen Abweichungen

$$q(\beta) := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (1.6)$$

minimiert. Für eine fest vorgegebenen Datensatz von Paaren  $(x_i, y_i)$  für  $i = 1, \dots, n$  ist der Wert dieser Summe abhängig von den Werten von  $\beta_0$  und  $\beta_1$  und kann deshalb durch Wahl geeigneter Werte von  $\beta_0$  und  $\beta_1$  minimiert werden. Da hierbei eine Summe von quadrierten Abweichungen zwischen Datenpunkten und Werten der Ausgleichsgerade minimiert wird, spricht man auch oft etwas ungenau von der *Methode der kleinsten Quadrate* (engl. *method of least squares*).

Abbildung 1.3 zeigt die vertikalen Abweichungen zwischen  $y_i$  und  $\beta_0 + \beta_1 x_i$  für  $i = 1, \dots, n$  des Beispieldatensatzes als orange Linien sowie die Summe ihrer Quadrate  $q(\beta)$  im Titel. Für die Parameterwerte  $\beta_0 = -6.2$  und  $\beta_1 = 1.7$  (vgl. Abbildung 1.2) nimmt diese Summe ihren kleinsten Wert an.



**Abbildung 1.3** Vertikale Abweichungen und Quadratsummen bei unterschiedlichen Parameterwerten.

Konkrete Formeln zur Bestimmung der Parameterwerte der Ausgleichsgerade stellt Theorem 1.1 bereit.

**Theorem 1.1** (Ausgleichsgerade). Für einen Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  hat die Ausgleichsgerade die Form

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \hat{\beta}_0 + \hat{\beta}_1 x, \quad (1.7)$$

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

$$\hat{\beta}_1 = \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (1.8)$$

◦

*Beweis.* Wir betrachten die Summe der quadrierten vertikalen Abweichungen der  $y_i$  von den Funktionswerten  $f(x_i)$  als Funktion von  $\beta_0$  und  $\beta_1$  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

$$q : \mathbb{R}^2 \rightarrow \mathbb{R}, (\beta_0, \beta_1) \mapsto q(\beta_0, \beta_1) := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2. \quad (1.9)$$

Um das Minimum dieser Funktion zu bestimmen, berechnen wir zunächst die partiellen Ableitungen hinsichtlich  $\beta_0$  und  $\beta_1$  und setzen diese gleich 0. Es ergibt sich zunächst:

$$\begin{aligned} \frac{\partial}{\partial \beta_0} q(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_0} \left( \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_0} (y_i - (\beta_0 + \beta_1 x_i))^2 \\ &= \sum_{i=1}^n 2(y_i - (\beta_0 + \beta_1 x_i)) \frac{\partial}{\partial \beta_0} (y_i - \beta_0 - \beta_1 x_i) \\ &= -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i). \end{aligned} \quad (1.10)$$

Weiterhin ergibt sich:

$$\begin{aligned} \frac{\partial}{\partial \beta_1} q(\beta_0, \beta_1) &= \frac{\partial}{\partial \beta_1} \left( \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} (y_i - (\beta_0 + \beta_1 x_i))^2 \\ &= \sum_{i=1}^n 2(y_i - (\beta_0 + \beta_1 x_i)) \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 x_i) \\ &= -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i. \end{aligned} \quad (1.11)$$

Nullsetzen beider partieller Ableitungen ergibt dann

$$\begin{aligned} \frac{\partial}{\partial \beta_0} q(\beta_0, \beta_1) = 0 \text{ und } \frac{\partial}{\partial \beta_1} q(\beta_0, \beta_1) = 0 \\ \Leftrightarrow -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \text{ und } -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \\ \Leftrightarrow \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \text{ und } \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{aligned} \quad (1.12)$$

und weiter

$$\begin{aligned} \sum_{i=1}^n y_i - \sum_{i=1}^n \beta_0 - \beta_1 \sum_{i=1}^n x_i &= 0 \text{ und } \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \beta_0 x_i - \beta_1 \sum_{i=1}^n x_i^2 = 0 \\ \Leftrightarrow \beta_0 n + \beta_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \text{ und } \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i . \end{aligned} \quad (1.13)$$

Das sich hier ergebende Gleichungssystem

$$\begin{aligned} \beta_0 n + \beta_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \end{aligned} \quad (1.14)$$

wird *System der Normalgleichungen* genannt und beschreibt die notwendige Bedingung für ein Minimum von  $q$ . Auflösen dieses Gleichungssystems nach  $\beta_0$  und  $\beta_1$  liefert dann die Werte  $\hat{\beta}_0$  und  $\hat{\beta}_1$  des Theorems.

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

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \Leftrightarrow \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} \Leftrightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (1.15)$$

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

$$\begin{aligned} \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \\ \Leftrightarrow (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \\ \Leftrightarrow \bar{y} \sum_{i=1}^n x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i \\ \Leftrightarrow -\hat{\beta}_1 \bar{x} \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i \\ \Leftrightarrow \hat{\beta}_1 \left( \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right) &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i . \end{aligned} \quad (1.16)$$

Wir halten nun fest, dass gilt

$$\begin{aligned} \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \bar{x} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n \left( \frac{1}{n} \sum_{i=1}^n x_i \right) \bar{x} \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 . \end{aligned} \quad (1.17)$$

Weiterhin halten wir fest, dass gilt

$$\begin{aligned}
 \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i - n\bar{y}\bar{x} + n\bar{y}\bar{x} \\
 &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i - \sum_{i=1}^n y_i \bar{x} + \sum_{i=1}^n \bar{y} \bar{x} \\
 &= \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \bar{x} - \sum_{i=1}^n \bar{y} x_i + \sum_{i=1}^n \bar{y} \bar{x} \\
 &= \sum_{i=1}^n (y_i x_i - y_i \bar{x} - \bar{y} x_i + \bar{y} \bar{x}) \\
 &= \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) .
 \end{aligned} \tag{1.18}$$

In der Fortsetzung von (1.16) ergibt sich dann

$$\begin{aligned}
 \hat{\beta}_1 \left( \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right) &= \sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i \\
 \Leftrightarrow \hat{\beta}_1 \left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) &= \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \\
 \Leftrightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
 \Leftrightarrow \hat{\beta}_1 &= \frac{c_{xy}}{s_x^2} .
 \end{aligned} \tag{1.19}$$

□

Theorem 1.1 besagt, dass die Parameterwerte, die für einen gegebenen Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  die Summe der quadrierten vertikalen Abweichungen für eine linear-affine Funktion minimieren, mithilfe der Stichprobenmittel der  $x_i$ - und  $y_i$ -Werte, der Stichprobenvarianz der  $x_i$ -Werte und der Stichprobenkovarianz der  $x_i$ - und  $y_i$ -Werte berechnet werden können.

Die Terminologie orientiert sich hier an den Begrifflichkeiten der deskriptiven Statistik, insbesondere werden die  $x_i$ -Werte häufig nicht als Realisationen von Zufallsvariablen verstanden, der Begriff der Stichprobe wird jedoch trotzdem verwendet. Aus der Anwendungsperspektive können nach Theorem 1.1 die Parameter der Ausgleichsgerade also mithilfe der bekannten Funktionen für die Auswertung deskriptiver Statistiken bestimmt werden. Folgender **R**-Code demonstriert dies.

```

# Einlesen des Beispieldatensatzes
fname      = '../01_Regression/Daten/Regression_Beispieldatensatz.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)
```

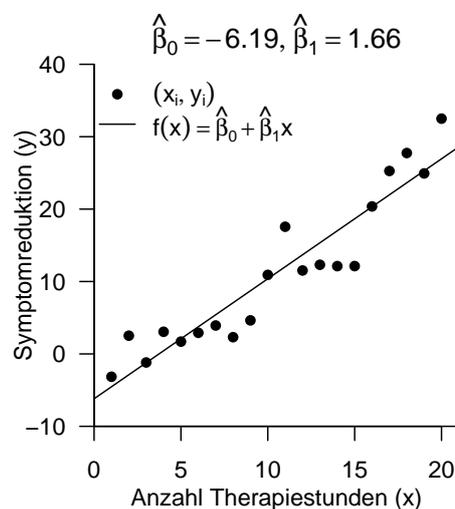
```
beta_0_hat: -6.194704
beta_1_hat: 1.657055
```

Eine typische Visualisierung der Ausgleichsgerade eines Datensatzes wie in Abbildung 1.4 wird durch folgenden **R**-Code implementiert.

```
# 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")
```



**Abbildung 1.4** Ausgleichsgerade für den Beispieldatensatz.

Die Idee, bei einem gegebenen Datensatz von  $(x_i, y_i)$  Paaren die Summe der quadrierten vertikalen Abweichungen zwischen einer Funktion der  $x_i$ -Werte und den  $y_i$ -Werten zu minimieren und so eine Funktion möglichst gut an eine Wertemenge anzupassen, ist nicht auf linear-affine Funktionen beschränkt. Beispielsweise lassen sich stattdessen auch Polynome höheren Grades betrachten.

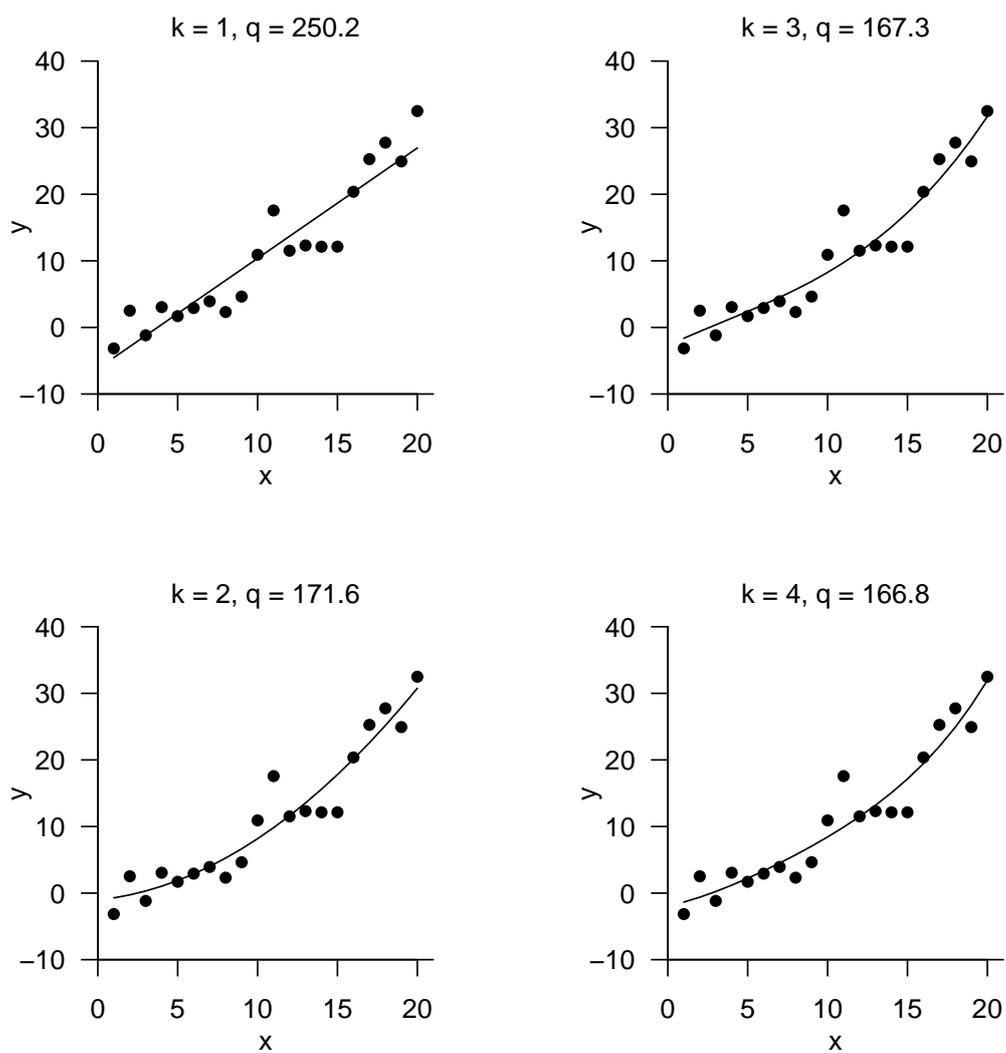


Abbildung 1.5 Ausgleichspolynome ersten bis vierten Grades für den Beispieldatensatz.

**Definition 1.3** (Polynom). Sei  $\beta := (\beta_0, \dots, \beta_k)^T \in \mathbb{R}^{k+1}$  ein  $(k+1)$ -dimensionaler Vektor reeller Zahlen. Dann ist ein Polynom  $k$ -ten Grades durch folgende Funktion gegeben:

$$f_\beta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_\beta(x) := \sum_{i=0}^k \beta_i x^i. \quad (1.20)$$

•

Folgende Definition verallgemeinert dann die Definition der Ausgleichsgerade auf Polynomfunktionen beliebigen Grades.

**Definition 1.4** (Ausgleichspolynom).  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  sei ein Datensatz. Weiterhin sei  $q(\beta)$  die Summe der quadrierten vertikalen Abweichungen der  $y_i$  von den Funktionswerten  $f_\beta(x_i)$ :

$$q : \mathbb{R}^{k+1} \rightarrow \mathbb{R}_{\geq 0}, \beta \mapsto q(\beta) := \sum_{i=1}^n (y_i - f_\beta(x_i))^2 = \sum_{i=1}^n \left( y_i - \sum_{j=0}^k \beta_j x_i^j \right)^2 \quad (1.21)$$

Dann heißt das Polynom  $k$ -ten Grades  $f_\beta$ , für das die Funktion  $q$  ihr Minimum annimmt, das *Ausgleichspolynom  $k$ -ten Grades* für den Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ .

•

Die Ausgleichsgerade ist damit das Ausgleichspolynom ersten Grades. Wir wollen den Begriff des Ausgleichspolynoms hier nicht weiter vertiefen und werden insbesondere die Parameterwerte  $\hat{\beta}_0, \dots, \hat{\beta}_k$ , für die die Funktion  $q$  ihr Minimum annimmt an späterer Stelle im Rahmen der Theorie des Allgemeinen Linearen Modells allgemein bestimmen. In Abbildung 1.5 visualisieren beispielhaft die Ausgleichspolynome ersten bis vierten Grades für den Beispieldatensatz wobei der Wert der Funktion  $q$  an der Minimumsstelle jeweils im Titel vermerkt ist.

## 1.2 Einfache lineare Regression

Eine Ausgleichsgerade erlaubt Aussagen über unbeobachtete Werte der abhängigen Variable. Allerdings erlaubt eine Ausgleichsgerade nur implizite Aussagen über die mit der Anpassung einer linear-affinen Funktion an einen Datensatz verbundene Unsicherheit. In der einfachen linearen Regression wird die Idee einer Ausgleichsgerade um eine probabilistische Komponente erweitert. Die einfache lineare Regression erlaubt damit insbesondere, im Sinne der Frequentistischen Inferenz Konfidenzintervalle für die Ausgleichsgeradenparameter anzugeben und Hypothesentests bezüglich der Ausgleichsgeradenparameter durchzuführen.

Wir wollen hier zunächst nur das Modell der einfachen linearen Regression und die auf ihm basierende Maximum-Likelihood-Schätzung der Ausgleichsgeradenparameter betrachten. Die Bewertung der mit dieser Schätzung verbundenen Unsicherheit sowie parameterzentrierte Hypothesentests behandeln wir dann an späterer Stelle zunächst im Kontext des Allgemeinen Linearen Modells. Wir beginnen mit folgender Definition.

**Definition 1.5** (Modell der einfachen linearen Regression). Es sei

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. f\u00fcr } i = 1, \dots, n, \quad (1.22)$$

wobei

- $y_i$  beobachtbare Zufallsvariablen sind, die Werte einer abh\u00e4ngigen Variable modellieren,
- $x_i \in \mathbb{R}$  fest vorgegebene sogenannte *Pr\u00e4diktorkwerte* oder *Regressorwerte* sind, die Werte einer unabh\u00e4ngigen Variable modellieren,
- $\beta_0, \beta_1 \in \mathbb{R}$  wahre, aber unbekannte \*Offset- und Steigungsparameterwerte\* sind und
- $\varepsilon_i \sim N(0, \sigma^2)$  unabh\u00e4ngig und identisch normalverteilte nicht-beobachtbare Zufallsvariablen mit wahrem, aber unbekanntem *Varianzparameter*  $\sigma^2 > 0$  sind, die Fehler- oder St\u00f6rvariablen modellieren.

Dann hei\u00dft Gleichung 1.22 *Modell der einfachen linearen Regression*.

•

Im Gegensatz zur Ausgleichsgerade treten im Modell der einfachen linearen Regression also explizit Zufallsvariablen auf. Speziell definiert das Modell der einfachen linearen Regression, wie  $n$  beobachtbare (abh\u00e4ngige) Zufallsvariablen  $y_i$  anhand der Werte  $x_i$  einer unabh\u00e4ngigen Variable, der Parameterwerte  $\beta_0$  und  $\beta_1$  sowie durch Addition der normalverteilten Fehlervariablen  $\varepsilon_i$  generiert werden. Das Modell hat dabei drei Parameter, den Offsetparameter  $\beta_0$ , den Steigungsparameter  $\beta_1$  und den Varianzparameter  $\sigma^2$  der normalverteilten Fehlervariablen. Addition der festen Werte  $\beta_0$  und  $\beta_1 x_i$  zu der normalverteilten Zufallsvariable  $\varepsilon_i$  impliziert dabei eine Normalverteilung von  $y_i$ . Dies ist die Aussage des folgenden Theorems.

**Theorem 1.2** (Datenverteilung der einfachen linearen Regression). *Das Modell der einfachen linearen Regression*

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. f\u00fcr } i = 1, \dots, n \quad (1.23)$$

l\u00e4sst sich mit  $\mu_i := \beta_0 + \beta_1 x_i$  \u00e4quivalent in der Form

$$y_i \sim N(\mu_i, \sigma^2) \text{ u.v. f\u00fcr } i = 1, \dots, n \quad (1.24)$$

schreiben.

o

*Beweis.* Wir zeigen die \u00c4quivalenz f\u00fcr ein  $i$ . Die Unabh\u00e4ngigkeit der  $y_i$  zeigen wir an sp\u00e4terer Stelle im Rahmen des Allgemeinen Linearen Modells. Die \u00c4quivalenz beider Modellformen f\u00fcr ein  $i$  folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen.

Allgemein gilt, gegeben eine Zufallsvariable  $\xi$ , dass sich die die Wahrscheinlichkeitsdichtefunktion der linear-affinen Transformation  $v := a\xi + b$  aus der Wahrscheinlichkeitsdichtefunktion von  $\xi$  ermitteln l\u00e4sst (siehe Einheit (7) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*):

$$v := a\xi + b \quad \Rightarrow \quad p_v(y) = \frac{1}{|a|} p_\xi\left(\frac{y-b}{a}\right). \quad (1.25)$$

Speziell gilt im vorliegenden Fall für  $\varepsilon_i \sim N(0, \sigma^2)$ , dass

$$y_i = f(\varepsilon_i) \text{ mit } f : \mathbb{R} \rightarrow \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + (\beta_0 + \beta_1 x_i). \quad (1.26)$$

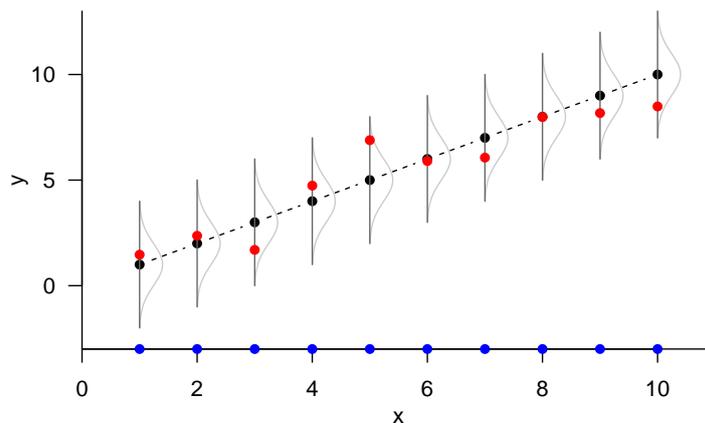
Mit dem Transformationstheorem für Wahrscheinlichkeitsdichtefunktionen bei linear-affinen Abbildungen in (??) folgt dann:

$$\begin{aligned} p(y_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left( \frac{y_i - \beta_0 - \beta_1 x_i}{1} \right) \\ &= N(y_i - \beta_0 - \beta_1 x_i; 0, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i - 0)^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - (\beta_0 + \beta_1 x_i))^2\right) \\ &= N(y_i; \beta_0 + \beta_1 x_i, \sigma^2). \end{aligned} \quad (1.27)$$

Definiert man  $\mu_i := \beta_0 + \beta_1 x_i$  für  $i = 1, \dots, n$ , ergibt sich die Aussage des Theorems.  $\square$

Theorem 1.2 besagt insbesondere, dass die Datenvariablen  $y_i$  univariat normalverteilte Zufallsvariablen sind, deren Erwartungswertparameter jeweils vom Wert der unabhängigen Variable  $x_i$  abhängen.

Abbildung 1.6 visualisiert das Modell und eine Realisation der einfachen linearen Regression für wahre, aber unbekannte, Parameterwerte  $\beta_0 := 0$ ,  $\beta_1 := 1$  und  $\sigma^2 := 1$ .



**Abbildung 1.6** Modell der einfachen linearen Regression für  $\beta_0 := 0$ ,  $\beta_1 := 1$  und  $\sigma^2 := 1$ .

Da es sich bei dem Modell der einfachen linearen Regression um ein parametrisches frequentistisches Modell handelt, können Schätzer für die Modellparameter mithilfe des Maximum-Likelihood-Prinzips gewonnen werden. Insbesondere stellt sich dabei heraus, dass die Maximum-Likelihood-Schätzer des Offset- und des Steigungsparameters mit den Werten der Ausgleichsgeradenparameter identisch sind. Dies ist eine der Aussagen folgenden Theorems. (Aus Gründen der Übersichtlichkeit verzichten wir hier auf die ML-Superskripte  $^{(ML)}$  und kennzeichnen Maximum-Likelihood-Schätzer durch das ‘Dach’-Zeichen  $(\hat{\cdot})$ .)

**Theorem 1.3** (Maximum-Likelihood Schätzung der einfachen linearen Regression). *Es sei*

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \quad (1.28)$$

das Modell der einfachen linearen Regression. Dann sind Maximum-Likelihood-Schätzer der Modellparameter  $\beta_0$ ,  $\beta_1$  und  $\sigma^2$  gegeben durch

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2}, \quad \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und} \quad \hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2. \quad (1.29)$$

◦

*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  $\beta_0$  und  $\beta_1$  die folgende Form hat:

$$\begin{aligned} L : \mathbb{R}^2 \rightarrow \mathbb{R}_{>0}, (\beta_0, \beta_1) &\mapsto L(\beta_0, \beta_1) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2\right). \end{aligned} \quad (1.30)$$

Für die Exponentialfunktion gilt: Wenn  $a < b \leq 0$  gilt, dann gilt  $\exp(a) < \exp(b)$ . Somit wird der Exponentialterm dieser Likelihood-Funktion maximal, wenn der Term

$$q := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \geq 0 \quad (1.31)$$

minimal und damit  $-q$  maximal wird. Im Rahmen des Beweises der Ausgleichsgeradenform haben wir aber schon gezeigt, dass der Term (1.31) für

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2} \text{ und } \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad (1.32)$$

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

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

$$L : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, \sigma^2 \mapsto L(\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2\right) \quad (1.33)$$

und die entsprechende Log-Likelihood-Funktion

$$\ell : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto \ell(\sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2. \quad (1.34)$$

In Analogie zu der Herleitung des ML-Schätzers für  $\sigma^2$  im Normalverteilungsmodell (vgl. Einheit (9) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*) ergibt sich unter Beachtung von

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (1.35)$$

dann hier

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2. \quad (1.36)$$

□

In der Anwendung ist die Maximum-Likelihood Schätzung der Parameter der einfachen linearen Regression also im Wesentlichen mit der Bestimmung der Ausgleichsgeradenparameter identisch, wie folgender **R**-Code zur Schätzung der Parameter für den Beispieldatensatz demonstriert.

```
# Einlesen des Beispieldatensatzes
fname      = '../01_Regression/Daten/Regression_Beispieldatensatz.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

# Parameteterschätzer
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) # \hat{\sigma}^2, Varianzparameter

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

```
beta_0_hat: -6.194704
beta_1_hat: 1.657055
sigsqr_hat: 3.536795
```

### 1.3 Literaturhinweise

Die Idee der Minimierung einer Summe von quadrierten Abweichungen bei der Anpassung einer Polynomfunktion an beobachtete Werte geht auf die Arbeiten von Legendre (1805) und Gauss (1809) im Kontext der Bestimmung von Planetenbahnen zurück. Eine historische Einordnung dazu gibt Stigler (1981). Der Begriff der Regression geht zurück auf Galton (1886). Stigler (1986) gibt dazu einen ausführlichen historischen Überblick.

### 1.4 Selbstkontrollfragen

1. Geben Sie die funktionale Form einer 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 Modells der einfachen linearen Regression in Bezug auf die Ausgleichsgerade.
9. Definieren Sie das Modell der einfachen linearen Regression.
10. Geben Sie das Theorem zur Datenverteilung 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.

## 2 Korrelation

Die Konzepte von Regression und Korrelation sind eng miteinander verwoben und dienen letztlich der Quantifizierung linear-affiner Abhängigkeiten zwischen unabhängigen und abhängigen Variablen. Ein zentrales Thema dieses Abschnittes sind damit die Äquivalenzen zwischen den beiden Konzepten.

Im Gegensatz zur Regression werden bei der Korrelation sowohl die abhängige als auch die unabhängige Variable als Zufallsvariablen modelliert, die probabilistische Komplexität der Korrelation ist also etwas höher als die der einfachen linearen Regression und des Allgemeinen Linearen Modells.

Um uns den Zusammenhängen von Regression und Korrelation zu nähern, erinnern wir Kapitel 2.1 zunächst an den Begriff und die Mechanik der Stichprobenkorrelation. In Kapitel 2.2 widmen wir uns anhand der  $R^2$ -Statistik dem Zusammenhang zwischen Stichprobenkorrelation und Ausgleichsgerade und geben ein erstes Beispiel für die Idee einer Varianzzerlegung. In Kapitel 2.3 schließlich diskutieren wir Aspekte zum Verhalten der (Stichproben-)Korrelation bei linear-affiner Transformation. Als Anwendungsbeispiel betrachten wir durchgängig den in Kapitel 1 eingeführten Beispieldatensatz zum Zusammenhang von Psychotherapiedauer und Symptomreduktion.

### 2.1 Grundlagen

Wir erinnern zunächst an den Begriff der Korrelation zweier Zufallsvariablen.

**Definition 2.1** (Korrelation). Die Korrelation\* zweier Zufallsvariablen  $\xi$  und  $v$  ist definiert als

$$\rho(\xi, v) := \frac{\mathbb{C}(\xi, v)}{\mathbb{S}(\xi)\mathbb{S}(v)} \quad (2.1)$$

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

•

Die Zahl  $\rho(\xi, v)$  wird auch Korrelationskoeffizient von  $\xi$  und  $v$  genannt. Wir haben bereits gesehen, dass  $-1 \leq \rho(\xi, v) \leq 1$  gilt und dass  $\xi$  und  $v$  unkorreliert genannt werden, wenn  $\rho(\xi, v) = 0$  ist. Weiterhin haben wir bereits gesehen, dass aus der Unabhängigkeit der Zufallsvariablen  $\xi$  und  $v$  immer die Unkorreliertheit von  $\xi$  und  $v$  folgt, dass aber umgekehrt im Allgemeinen aus der Unkorreliertheit von  $\xi$  und  $v$  nicht die Unabhängigkeit von  $\xi$  und  $v$  folgt.

Liegen von zwei Zufallsvariablen Realisationen  $\{(x_i, y_i)\}_{i=1, \dots, n}$  als ein bivariater Datensatz vor, so kann man anhand folgender Definition die realisationsspezifische Stichprobenkorrelation bestimmen.

**Definition 2.2** (Stichprobenkorrelation).  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  sei ein Datensatz. Weiterhin seien:

- die Stichprobenmittel der  $x_i$  und  $y_i$  definiert als

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i \text{ und } \bar{y} := \frac{1}{n} \sum_{i=1}^n y_i, \quad (2.2)$$

- die Stichprobenstandardabweichungen  $x_i$  und  $y_i$  definiert als

$$s_x := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \text{ und } s_y := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.3)$$

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

$$c_{xy} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (2.4)$$

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

$$r_{xy} := \frac{c_{xy}}{s_x s_y} \quad (2.5)$$

und wird auch *Stichprobenkorrelationskoeffizient* genannt.

•

Folgender **R**-Code wertet die Stichprobenkorrelation des Beispieldatensatzes aus.

```
# Laden des Beispieldatensatzes
fname = "../02_Korrelation/Daten/Korrelation_Beispieldatensatz.csv" # Dateipfad generieren
D = read.table(fname, sep = ",", header = TRUE) # Beispieldatensatz als Dataframe laden
x_i = D$x_i # x_i Werte
y_i = D$y_i # y_i Werte
n = length(x_i) # n

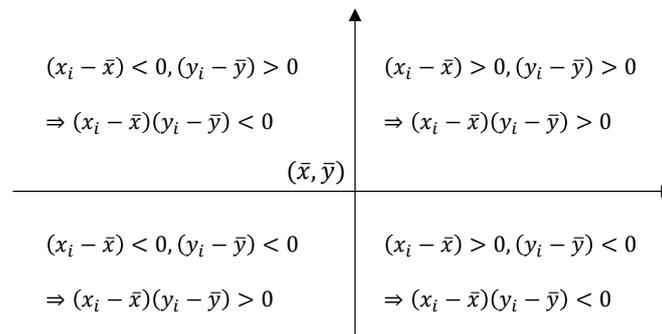
# "manuelle" Berechnung der Stichprobenkorrelation
x_bar = (1/n)*sum(x_i) # \bar{x}
y_bar = (1/n)*sum(y_i) # \bar{y}
s_x = sqrt(1/(n-1)*sum((x_i - x_bar)^2)) # s_x
s_y = sqrt(1/(n-1)*sum((y_i - y_bar)^2)) # s_y
c_xy = 1/(n-1) * sum((x_i - x_bar) * (y_i - y_bar)) # c_{xy}
r_xy = c_xy/(s_x * s_y) # r_{xy}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

```
[1] 0.9378162
```

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

```
[1] 0.9378162
```

Im Beispieldatensatz sind die Anzahl der Therapiestunden und die Symptomreduktion also mit  $r_{xy} = 0.94$  hoch korreliert. Allgemein spricht man bei absoluten Werten von  $r_{xy}$  größer als etwa 0.70 von hoher Korrelation, bei absoluten Werten von  $r_{xy}$  zwischen etwa 0.30 und 0.70 von mittlerer Korrelation und bei absoluten Werten von  $r_{xy}$  zwischen 0.00 und 0.30 von niedriger Korrelation. Eine niedrige Korrelation zweier Variablen bedeutet



**Abbildung 2.1** Mechanik der Stichprobenkorrelationssummenterme.

aber nicht zwangsläufig, dass diese Korrelation irrelevant ist (man denke an Gesundheitsrisikofaktoren), und ebenso kann auch eine hohe Korrelation zweier Variablen trivial sein (man denke an die Korrelation von Körpergröße und Schuhgröße).

Da die Stichprobenkorrelation lediglich die auf das Intervall  $[-1, 1]$  normalisierte Stichprobenkovarianz  $c_{xy}$  ist, wird die Höhe der Stichprobenkorrelation und insbesondere ihr Vorzeichen entscheidend durch die Werte Stichprobenkovarianzsummenterme  $(x_i - \bar{x})(y_i - \bar{y})$  bestimmt. Dabei ist es entscheidend, wie häufig über die Datenpaare  $(x_i, y_i)$  hinweg die  $x_i$  und  $y_i$  gleichartig oder entgegengesetzt von ihren jeweiligen Stichprobenmitteln abweichen. Dies ist schematisch in [Abbildung 2.1](#) dargestellt.

Bei häufiger richtungsgleicher Abweichung von ihren jeweiligen Mittelwerten, sowohl in positiver als auch in negativer Richtung, ergibt das Produkt  $(x_i - \bar{x})(y_i - \bar{y})$  eine positive Zahl, trägt also zu einer positiven Stichprobenkovarianz bei. Bei häufiger entgegengesetzter Abweichung von ihren jeweiligen Mittelwerten ergibt das Produkt  $(x_i - \bar{x})(y_i - \bar{y})$  häufig eine negative Zahl, trägt also zu einer negativen Stichprobenkovarianz bei. Kommen sowohl richtungsgleiche als auch entgegengesetzte Abweichungen der  $x_i$  und  $y_i$  häufig vor, so gleichen sich positive und negative Beiträge zur Stichprobenkovarianzsumme eher aus und es resultiert eine geringe Stichprobenkovarianz bzw. ein Stichprobenkorrelationskoeffizient nahe Null. [Abbildung 2.2](#) zeigt bivariate Datensätze von jeweils  $n = 30$  Datenpunkten zusammen mit ihren jeweiligen Stichprobenkorrelationskoeffizienten.

## 2.2 Korrelation und Bestimmtheitsmaß

Das sogenannte Bestimmtheitsmaß  $R^2$  ist eine beliebte, häufig berichtete Statistik zur Beschreibung der Stärke des Zusammenhangs der Werte einer unabhängigen und einer abhängigen Variable. Numerisch handelt es sich im Falle von zwei Variablen bei  $R^2$  lediglich um den quadrierten Stichprobenkorrelationskoeffizienten. Ist die Stichprobenkorrelation zum Beispiel  $r_{xy} = 0.5$ , dann ist  $R^2 = 0.5^2 = 0.25$ . Ist die Stichprobenkorrelation dagegen  $r_{xy} = -0.5$ , dann gilt analog  $R^2 = (-0.5)^2 = 0.25$ . An diesen Beispielen erkennt man, dass  $R^2$  weniger Information über die Rohdaten enthält als  $r_{xy}$ , da das Vorzeichen und damit die Richtung des Zusammenhangs wegfällt. *Per se* ist die Angabe von  $R^2$  anstelle von  $r_{xy}$  als Deskriptivstatistik zur Beschreibung der Zusammenhangsstärke der Werte von unabhängiger und abhängiger Variable daher ohne Vorteil.

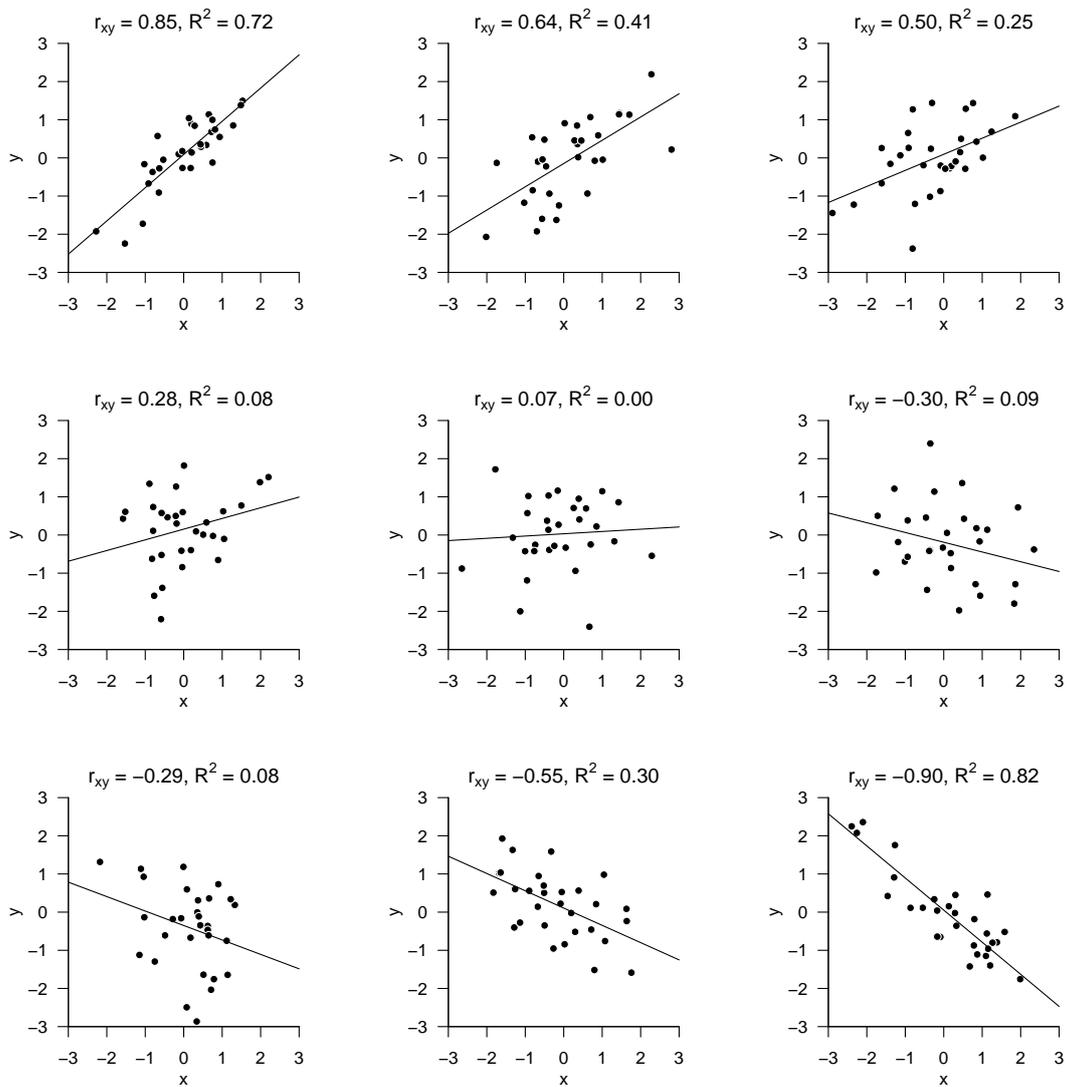


Abbildung 2.2 Beispiele bivariater Datensätze und ihrer Stichprobenkorrelationen.

Wir wollen hier trotzdem etwas genauer auf  $\mathbb{R}^2$  eingehen, da ein tieferes Verständnis von  $\mathbb{R}^2$  einerseits den Einstieg in das Konzept der Varianzzerlegung erlaubt und andererseits die Zusammenhänge zwischen den Konzepten der Ausgleichsgerade und der Stichprobenkorrelation weiter verdeutlicht. Wir erweitern dazu zunächst die Beschreibung der Ausgleichsgerade aus Kapitel 1.1 durch die Begriffe der erklärten Werte und der Residuen.

**Definition 2.3** (Erklärte Werte und Residuen einer Ausgleichsgerade.). Gegeben seien ein Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  und die zu diesem Datensatz gehörende Ausgleichsgerade

$$f_{\hat{\beta}} : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_{\hat{\beta}}(x) := \hat{\beta}_0 + \hat{\beta}_1 x. \quad (2.6)$$

Dann werden für  $i = 1, \dots, n$

$$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (2.7)$$

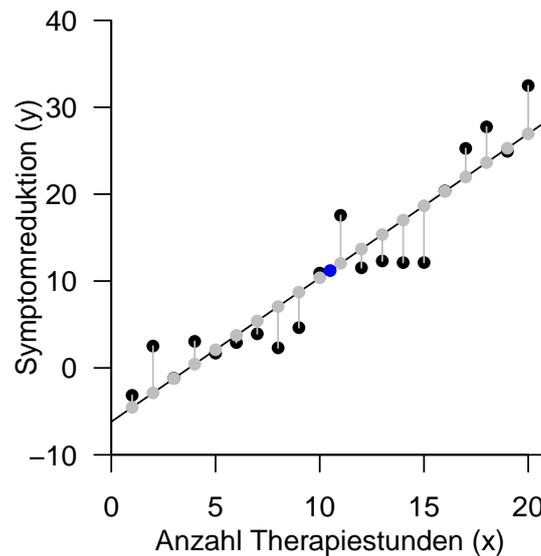
die durch die Ausgleichsgerade *erklärten Werte* genannt und

$$\hat{\varepsilon}_i := y_i - \hat{y}_i \quad (2.8)$$

die *Residuen* der Ausgleichsgerade genannt.

•

Etwas allgemeiner formuliert sind die *erklärten Werte* damit die Datenvorhersage des Modells basierend auf den geschätzten Parameterwerten, während die Residuen die Differenzen zwischen den geschätzten Datenvorhersagen und den beobachteten Datenwerten bezeichnen. Abbildung 2.3 verdeutlicht diese Begriffe am Beispiel der Ausgleichsgerade des Beispieldatensatzes.



**Abbildung 2.3** Ausgleichsgerade (schwarze Linie), erklärte Werte (graue Punkte) und Residuen (graue Linien) für den Beispieldatensatz.

Mithilfe der Begriffe der erklärten Werte und Residuen lässt sich nun folgende Quadratsummenzerlegung für einen bivariaten Datensatz aus abhängiger und unabhängiger Variable angeben.

**Theorem 2.1** (Quadratsummenzerlegung bei Ausgleichsgerade). *Für einen Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  und seine zugehörige Ausgleichsgerade  $f_{\hat{\beta}}$  seien*

$$\bar{y} := \frac{1}{n} \sum_{i=1}^n y_i \text{ und } \hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \text{ für } i = 1, \dots, n \quad (2.9)$$

das Stichprobenmittel der  $y_i$ -Werte bzw. die durch die Ausgleichsgerade erklärten Werte. Weiterhin seien

- die totale Quadratsumme (*total sum of squares*) definiert als

$$\text{SQT} := \sum_{i=1}^n (y_i - \bar{y})^2, \quad (2.10)$$

- die erklärte Quadratsumme (*explained sum of squares*) definiert als

$$\text{SQE} := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \text{ und} \quad (2.11)$$

- die residuelle Quadratsumme (*residual sum of squares*) definiert als

$$\text{SQR} := \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.12)$$

Dann gilt:

$$\text{SQT} = \text{SQE} + \text{SQR}. \quad (2.13)$$

◦

*Beweis.* Wir schreiben zunächst die totale Quadratsumme aus:

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n ((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2 \\ &= \sum_{i=1}^n ((y_i - \hat{y}_i)^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2) \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \text{SQE} + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \text{SQR} \\ &= \text{SQE} + \text{SQR} \end{aligned} \quad (2.14)$$

Dabei ergibt sich die letzte Gleichung mit

$$\bar{\hat{y}} := \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y} \quad (2.15)$$

und damit auch

$$\bar{\hat{y}} = \bar{y} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \Leftrightarrow \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i \Leftrightarrow \bar{y} \sum_{i=1}^n \hat{y}_i = \bar{y} \sum_{i=1}^n y_i \quad (2.16)$$

sowie

$$\bar{\hat{y}} = \bar{y} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \Leftrightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i \Leftrightarrow \sum_{i=1}^n y_i \hat{y}_i = \sum_{i=1}^n \hat{y}_i \hat{y}_i \quad (2.17)$$

aus

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n (y_i \hat{y}_i - y_i \bar{y} - \hat{y}_i \hat{y}_i + \hat{y}_i \bar{y}) \\ &= \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n y_i \bar{y} - \sum_{i=1}^n \hat{y}_i \hat{y}_i + \sum_{i=1}^n \hat{y}_i \bar{y} \\ &= \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n \hat{y}_i \hat{y}_i + \bar{y} \sum_{i=1}^n \hat{y}_i - \bar{y} \sum_{i=1}^n y_i \\ &= 0 + 0 \\ &= 0. \end{aligned} \quad (2.18)$$

□

Die Begriffsbildungen von Theorem 2.1 erklären sich intuitiv wie folgt:

- SQT repräsentiert die Gesamtvariabilität der  $y_i$ -Werte um ihren Mittelwert  $\bar{y}$  im Sinne eines unnormalisierten Varianzterms.
- SQE repräsentiert die Variabilität der erklärten Werte  $\hat{y}_i$  um ihren Mittelwert. Große Werte von SQE repräsentieren damit eine große absolute Steigung der  $y_i$  mit den  $x_i$  und kleine Werte von SQE repräsentieren eine kleine absolute Steigung der  $y_i$  mit den  $x_i$ . SQE ist somit ein Maß für die Stärke des linearen Zusammenhangs der  $x_i$ -Werte mit den  $y_i$ -Werte.
- SQR ist die Summe der quadrierten Residuen, denn es gilt:

$$\text{SQR} := \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - f_{\hat{\beta}}(x_i))^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2. \quad (2.19)$$

Große Werte von SQR repräsentieren damit große Abweichungen der erklärten von den beobachteten  $y_i$ -Werten und kleine Werte von SQR repräsentieren geringe Abweichungen der erklärten von den beobachteten  $y_i$ -Werten. SQR ist also ein Maß für die Güte der Beschreibung der Datenmenge durch die Ausgleichsgerade.

Die zentrale Aussage des Theorem 2.1 ist nun, dass sich die Gesamtstreuung der  $y_i$ -Werte um ihren Mittelwert  $\bar{y}$  gerade aus der Summe der Stärke des linearen Zusammenhangs der  $x_i$ - und  $y_i$ -Werte (also des “deterministischen Einflusses” der  $x_i$  auf die  $y_i$ ) sowie den Abweichungen von diesem linearen Zusammenhang (also dem “Rauschen”) zusammensetzt.

Obwohl es sich bei SQT formal nicht um ein Varianzmaß handelt, spricht man in diesem Zusammenhang auch oft von einer Varianzzerlegung in erklärte Varianz und Residualvarianz. Dieses Motiv ist ein zentraler Aspekt des Allgemeinen Linearen Modells und wird in späteren Kapiteln erneut aufgegriffen werden. Für den Moment erlaubt Theorem 2.1 nun folgende Definition des Bestimmtheitsmaßes  $R^2$ .

**Definition 2.4** (Bestimmtheitsmaß  $R^2$ ). Für einen Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  und seine zugehörige Ausgleichsgerade  $f_{\hat{\beta}}$  sowie die zugehörige erklärte Quadratsumme SQE und totale Quadratsumme SQT heißt

$$R^2 := \frac{\text{SQE}}{\text{SQT}} \quad (2.20)$$

*Bestimmtheitsmaß* oder *Determinationskoeffizient*.

•

Ausgehend von dieser Definition und dem Theorem zur Quadratsummenzerlegung bei Ausgleichsgerade lässt sich  $R^2$  alternativ auch in Abhängigkeit der residuellen Quadratsumme SQR ausdrücken:

$$R^2 := \frac{\text{SQE}}{\text{SQT}} = \frac{\text{SQT} - \text{SQR}}{\text{SQT}} = \frac{\text{SQT}}{\text{SQT}} - \frac{\text{SQR}}{\text{SQT}} = 1 - \frac{\text{SQR}}{\text{SQT}}. \quad (2.21)$$

Folgendes Theorem liefert nun den oben erwähnten Zusammenhang zwischen dem Bestimmtheitsmaß und der Stichprobenkorrelation.

**Theorem 2.2** (Stichprobenkorrelation und Bestimmtheitsmaß). Für einen Datensatz  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$  sei  $R^2$  das Bestimmtheitsmaß und  $r_{xy}$  sei die Stichprobenkorrelation. Dann gilt:

$$R^2 = r_{xy}^2. \quad (2.22)$$

◦

*Beweis.* Wir halten zunächst fest, dass mit

$$\bar{\hat{y}} := \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y} \quad (2.23)$$

folgt, dass

$$\begin{aligned} \text{SQE} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \\ &= \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})^2 \\ &= \sum_{i=1}^n (\hat{\beta}_1 (x_i - \bar{x}))^2 \\ &= \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned} \quad (2.24)$$

Damit ergibt sich dann

$$\begin{aligned}
 R^2 &= \frac{\text{SQE}}{\text{SQT}} \\
 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
 &= \hat{\beta}_1^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
 &= \frac{c_{xy}^2 \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}{s_x^4 \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \\
 &= \frac{c_{xy}^2 s_x^2}{s_x^4 s_y^2} \\
 &= \frac{c_{xy}^2}{s_x^2 s_y^2} \\
 &= \left( \frac{c_{xy}}{s_x s_y} \right)^2 \\
 &= r_{xy}^2 .
 \end{aligned} \tag{2.25}$$

□

Man beachte, dass mit  $-1 \leq r_{xy} \leq 1$  aus Theorem 2.2 direkt folgt, dass  $0 \leq R^2 \leq 1$ . Nach Definition 2.4 gilt  $R^2 = 0$  genau dann, wenn  $\text{SQE} = 0$  (oder  $\text{SQR} = \text{SQT}$ ) ist.  $R^2 = 0$  bedeutet also, dass die erklärte Datenvariabilität durch die Ausgleichsgerade gleich Null ist und beschreibt damit den Fall einer denkbar schlechten Datenerklärung durch die Ausgleichsgerade.

Andererseits gilt  $R^2 = 1$  genau dann, wenn  $\text{SQE} = \text{SQT}$  (oder  $\text{SQR} = 0$ ) ist.  $R^2 = 1$  bedeutet also, dass die Gesamtstreuung gleich der durch die Ausgleichsgerade erklärten Streuung ist und beschreibt den Fall, dass sämtliche Datenvariabilität durch die Ausgleichsgerade erklärt werden kann. Man sagt deshalb auch oft etwas ungenau, dass  $R^2$  den Anteil der durch die Ausgleichsgerade erklärten Varianz an der Gesamtvarianz der Daten repräsentiert.

Neben der jeweiligen Stichprobenkorrelation ist in Abbildung 2.2 auch jeweils das Bestimmtheitsmaß für die bivariaten Beispieldatensätze aufgeführt.

### 2.3 Korrelation und linear-affine Abhängigkeit

Die Tatsache, dass stochastische Unabhängigkeit zwar Unkorreliertheit impliziert, umgekehrt die Unkorreliertheit zweier Zufallsvariablen aber nicht ihre stochastische Unabhängigkeit impliziert, deutet daraufhin, dass die Korrelation zweier Zufallsvariablen nur bestimmte Formen der Abhängigkeit zwischen Variablen misst. Abbildung 2.4 verdeutlicht dies anhand dreier Simulationsbeispiele.

Abbildung 2.4 A zeigt eine Realisation des Modells

$$y_i = x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, 1) \text{ für } i = 1, \dots, n . \tag{2.26}$$

Die Stichprobenkorrelationen des Datensatzes  $\{(x_i, y_i)\}_{i=1}^n$  ergibt sich hier zu  $r_{xy} = 0.90$ . Es besteht eine klare Abhängigkeit des Wertes der Realisation  $y_i$  vom Wert  $x_i$ : Je höher der Wert von  $x_i$ , desto höher der Erwartungswert für den Wert von  $y_i$ .

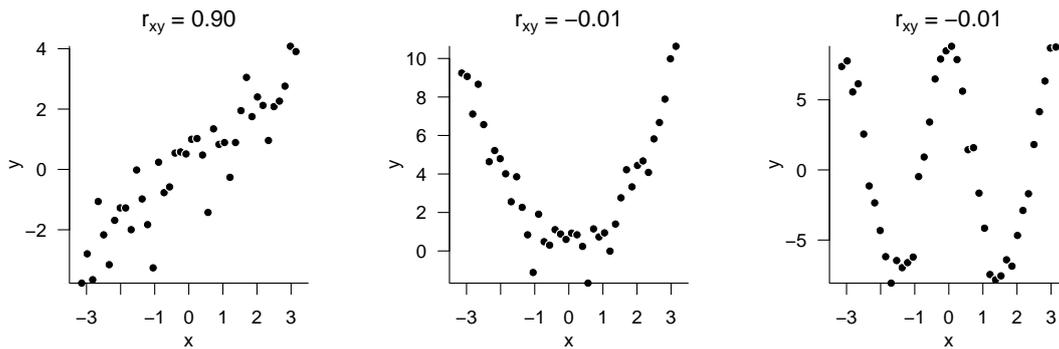


Abbildung 2.4 Korrelation und Abhängigkeit.

Abbildung 2.4 B zeigt eine Realisation des Modells

$$y_i = x_i^2 + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, 1) \text{ für } i = 1, \dots, n . \tag{2.27}$$

Die Stichprobenkorrelationen des Datensatzes  $\{(x_i, y_i)\}_{i=1}^n$  ergibt sich hier zu  $r_{xy} = -0.01$ . Die Stichprobenkorrelation ist also minimal. Es besteht aber auch hier eine klare Abhängigkeit des Wertes der Realisation  $y_i$  vom Wert  $x_i$ : Je höher oder je niedriger der Wert von  $x_i$ , desto höher der Erwartungswert für den Wert von  $y_i$ , es besteht ein quadratischer Zusammenhang.

Abbildung 2.4 C zeigt eine Realisation des Modells

$$y_i = 8 \cos(2x_i) + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, 1) \text{ für } i = 1, \dots, n . \tag{2.28}$$

Auch hier ergibt sich das Bild einer Abhängigkeit des Wertes der Realisation  $y_i$  vom Wert  $x_i$ , in diesem Fall im Sinne einer zyklischen Abhängigkeit, die Stichprobenkorrelation ist aber mit  $r_{xy} = -0.01$  wiederum minimal. Diese Simulationsbeispiele belegen also intuitiv, dass die Stichprobenkorrelation verschwindend gering sein kann, auch wenn klare Abhängigkeiten zwischen zwei Variablen bestehen.

Als allgemeines Maß für die Abhängigkeit zweier Variablen ist die Korrelation also ungeeignet. Diesen Umstand zugrunde liegt die Tatsache, dass die Korrelation lediglich ein Maß für den linear-affinen Zusammenhang zweier Zufallsvariablen, nicht aber für ihre stochastische Abhängigkeit ist. Wir formalisieren und präzisieren diese Aussage in folgenden Theorem.

**Theorem 2.3** (Korrelation und linear-affine Abhängigkeit).  $\xi$  und  $v$  seien zwei Zufallsvariablen mit positiver Varianz. Dann gilt: Ihre Korrelation beträgt

$$\rho(\xi, v) = 1 \text{ oder } \rho(\xi, v) = -1 \tag{2.29}$$

genau dann, wenn eine lineare-affine Abhängigkeit der folgenden Form zwischen  $\xi$  und  $v$  besteht:

$$v = \beta_0 + \beta_1 \xi \text{ mit } \beta_0, \beta_1 \in \mathbb{R} . \tag{2.30}$$

◦

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

$$\mathbb{E}(v) = \beta_0 + \beta_1 \mathbb{E}(\xi) \text{ und } \mathbb{V}(v) = \beta_1^2 \mathbb{V}(\xi) \text{ und somit } \mathbb{S}(v) = \pm \beta_1 \mathbb{S}(\xi) . \quad (2.31)$$

Wegen  $\mathbb{V}(\xi) > 0$  und  $\mathbb{V}(v) > 0$  gilt damit  $\beta_1 \neq 0$ . Weiterhin gilt:

$$\begin{aligned} v - \mathbb{E}(v) &= \beta_0 + \beta_1 \xi - \mathbb{E}(v) \\ &= \beta_0 + \beta_1 \xi - \beta_0 - \beta_1 \mathbb{E}(\xi) \\ &= \beta_1 \xi - \beta_1 \mathbb{E}(\xi) \\ &= \beta_1 (\xi - \mathbb{E}(\xi)) . \end{aligned} \quad (2.32)$$

Für die Kovarianz von  $\xi$  und  $v$  ergibt sich also:

$$\begin{aligned} \mathbb{C}(\xi, v) &= \mathbb{E}((v - \mathbb{E}(v))(\xi - \mathbb{E}(\xi))) \\ &= \mathbb{E}(\beta_1 (\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))) \\ &= \beta_1 \mathbb{E}((\xi - \mathbb{E}(\xi))^2) \\ &= \beta_1 \mathbb{V}(\xi) . \end{aligned} \quad (2.33)$$

Damit ergibt für die Korrelation von  $\xi$  und  $v$ :

$$\rho(\xi, v) = \frac{\mathbb{C}(\xi, v)}{\mathbb{S}(\xi)\mathbb{S}(v)} = \pm \frac{\beta_1 \mathbb{V}(\xi)}{\mathbb{S}(\xi)\beta_1 \mathbb{S}(\xi)} = \pm \frac{\beta_1 \mathbb{V}(\xi)}{\beta_1 \mathbb{V}(\xi)} = \pm 1. \quad (2.34)$$

□

Die *Korrelation* zweier Zufallsvariablen wird also genau dann maximal, wenn zwischen den beiden Zufallsvariablen ein linear-affiner Zusammenhang besteht. Dabei impliziert die linear-affine Abhängigkeit der  $v$  von den  $\xi$  auch immer die linear-affine Abhängigkeit der  $\xi$  von den  $v$ , denn

$$v = \beta_0 + \beta_1 \xi \Leftrightarrow -\beta_0 + v = \beta_1 \xi \Leftrightarrow \xi = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} v \Leftrightarrow \xi = \tilde{\beta}_0 + \tilde{\beta}_1 v \quad (2.35)$$

mit

$$\tilde{\beta}_0 = -\frac{\beta_0}{\beta_1} \text{ und } \tilde{\beta}_1 = \frac{1}{\beta_1} . \quad (2.36)$$

Man kann auch Aussagen zum Verhalten der *Stichprobenkorrelation* bei linear-affiner Transformation treffen, die zudem den Vorteil des Stichprobenkorrelationskoeffizienten gegenüber der Stichprobenkovarianz als Zusammenhangsmaß deutlich machen. Dieser Vorteil besteht darin, dass der absolute Wert des Stichprobenkorrelationskoeffizienten bei linear-affiner Transformation der zugrundeliegenden Wertemenge gleich bleibt, wohingegen die Stichprobenkovarianz ihren Wert je nach gewähltem Maßstab ändert. Man sagt deshalb auch, dass der Stichprobenkorrelationkoeffizient maßstabsunabhängig ist. Dies ist die zentrale Aussage folgenden Theorems.

**Theorem 2.4** (Stichprobenkorrelation bei linear-affiner Transformation). *Für einen Datensatz  $\{(x_i, y_i)\}_{i=1, \dots, n} \subset \mathbb{R}^2$  sei  $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1, \dots, n} \subset \mathbb{R}^2$  eine linear-affin transformierte Wertemenge mit*

$$(\tilde{x}_i, \tilde{y}_i) = (a_x x_i + b_x, a_y y_i + b_y), a_x, a_y \neq 0 . \quad (2.37)$$

Dann gilt:

$$|r_{\tilde{x}\tilde{y}}| = |r_{xy}| . \quad (2.38)$$

◦

*Beweis.* Es gilt

$$\begin{aligned}
 r_{\tilde{x}\tilde{y}} &:= \frac{\frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_i - \tilde{\bar{x}})(\tilde{y}_i - \tilde{\bar{y}})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_i - \tilde{\bar{x}})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \tilde{\bar{y}})^2}} \\
 &= \frac{\sum_{i=1}^n (a_x x_i + b_x - (a_x \bar{x} + b_x))(a_y y_i + b_y - (a_y \bar{y} + b_y))}{\sqrt{\sum_{i=1}^n (a_x x_i + b_x - (a_x \bar{x} + b_x))^2} \sqrt{\sum_{i=1}^n (a_y y_i + b_y - (a_y \bar{y} + b_y))^2}} \\
 &= \frac{a_x a_y \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{a_x^2 \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{a_y^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \\
 &= \frac{a_x a_y}{|a_x| |a_y|} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\
 &= \frac{a_x a_y}{|a_x| |a_y|} \frac{c_{xy}}{s_x s_y} \\
 &= \frac{a_x a_y}{|a_x| |a_y|} r_{xy}.
 \end{aligned} \tag{2.39}$$

Nach Durchspielen aller möglichen Vorzeichenfälle ergibt sich:

$$|r_{\tilde{x}\tilde{y}}| = |r_{xy}|. \tag{2.40}$$

□

## 2.4 Literaturhinweise

Der Begriff der Korrelation erscheint, allerdings basierend auf früheren Arbeiten zum Beispiel von Bravais (1844), zunächst bei Galton (1890) (vgl. Stigler (1986)) und wird unter anderem durch die Arbeiten von Pearson (1895), Pearson (1896), Pearson (1900), Pearson (1901) im Kontext multivariater Normalverteilungen weiter ausgearbeitet. Eine frühe Studie zum Verhältnis von Korrelation und Kausalität ist Wright (1921).

## 2.5 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. Geben Sie die Definitionen von erklärten Werten und Residuen einer Ausgleichsgerade wieder.
5. Geben Sie das Theorem zur Quadratsummenzerlegung bei einer Ausgleichsgerade wieder.
6. Erläutern Sie die intuitiven Bedeutungen von SQT, SQE und SQR.
7. Geben Sie die Definition des Bestimmtheitsmaßes  $R^2$  wieder.
8. Geben Sie das Theorem zum Zusammenhang von Stichprobenkorrelation und Bestimmtheitsmaß wieder.
9. Erläutern Sie die Bedeutung von hohen und niedrigen Werten von  $R^2$  im Lichte der Ausgleichsgerade.
10. Geben Sie das Theorem zum Zusammenhang von Korrelation und linear-affiner Abhängigkeit wieder.
11. Geben Sie das Theorem zur Stichprobenkorrelation bei linear-affinen Transformationen wieder.
12. Erläutern Sie das Theorem zur Stichprobenkorrelation bei linear-affinen Transformationen.

# 3 Matrizen

Matrizen sind die Worte in der Sprache der modernen Datenanalyse. Ein Verständnis moderner datenanalytischer Verfahren und ihrer Implementation ist ohne ein Grundverständnis des Matrixbegriffs und ein Wissen um die grundlegenden Matrixoperationen nicht möglich. Matrizen können dabei sehr unterschiedliche Rollen spielen. So können Matrizen zum Beispiel Daten, experimentelle Designs und Modellparameter repräsentieren. Im Kontext der Linearen Algebra dienen Matrizen zur Repräsentation linearer Abbildungen und von Vektorräumen, hier werden Vektoren dann als spezielle Matrizen aufgefasst.

In diesem Kapitel geben wir eine Einführung zum Umgang mit Matrizen, wobei wir im Wesentlichen auf abstrakte Begrifflichkeiten der Linearen Algebra verzichten. Wir führen zunächst den Matrixbegriff ein und diskutieren dann mit der Matrixaddition, der Matrixsubtraktion, der Skalarmultiplikation und der Matrixtransposition erste grundlegende Matrixoperationen (Kapitel 3.1 und Kapitel 3.2). Wir führen dann die zentralen Begriffe der Matrixmultiplikation und der Matrixinversion ein (Kapitel 3.3 und Kapitel 3.4). Mit der Matrixdeterminante diskutieren wir dann in Kapitel 3.5 eine erste Maßzahl zur Beschreibung von Matrizen. Wir schließen in Kapitel 3.6 mit einer Übersicht zu besonders häufig auftretenden Matrizen.

## 3.1 Definition

Wir beginnen mit der Definition einer Matrix.

**Definition 3.1.** Eine Matrix ist eine rechteckige Anordnung von Zahlen, die wie folgt bezeichnet wird

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}. \tag{3.1}$$

•

Matrizen bestehen aus *Zeilen (rows)* und *Spalten (columns)*. Die Matrixeinträge  $a_{ij}$  werden mit einem *Zeilenindex*  $i$  und einem *Spaltenindex*  $j$  bezeichnet. Zum Beispiel gilt für

$$A := \begin{pmatrix} 2 & 7 & 5 & 2 \\ 8 & 2 & 5 & 6 \\ 6 & 4 & 0 & 9 \\ 9 & 2 & 1 & 2 \end{pmatrix}, \tag{3.2}$$

dass  $a_{32} = 4$ , da der zweite Eintrag in der dritten Zeile (oder alternativ, der dritte Eintrag in der zweiten Spalte) 4 ist. 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  $a_{ij} \in \mathbb{R}$  für alle  $i = 1, \dots, n$  und  $j = 1, \dots, m$ . Wir nennen Matrizen mit reellen Einträge *reelle Matrizen* und bezeichnen die Menge der reellen Matrizen mit  $n$  Zeilen und  $m$  Spalten mit  $\mathbb{R}^{n \times m}$ . An dem Ausdruck

$$A \in \mathbb{R}^{n \times m} \quad (3.3)$$

können wir also ablesen, dass  $A$  eine reelle Matrix mit  $n$  Zeilen und  $m$  Spalten ist. Wir identifizieren dabei die Menge  $\mathbb{R}^{1 \times 1}$  mit der Menge  $\mathbb{R}$  und die Menge  $\mathbb{R}^{n \times 1}$  mit der Menge  $\mathbb{R}^n$ . Reelle Matrizen mit einer Spalte und  $n$  Zeilen entsprechen also  $n$ -dimensionalen reellen Vektoren und reelle Matrizen mit einer Spalte und einer Zeile entsprechen reellen Zahlen.

### Definition von Matrizen in R

In **R** werden Matrizen definiert, indem Vektoren mithilfe der Funktion `matrix()` in die Repräsentation einer mathematischen Matrix transformiert werden. Die Einträge eines Vektors werden dabei anhand der spezifizierten Zeilenanzahl `nrow` abhängig von ihrer Gesamtanzahl auf die Matrix verteilt. Wollen wir beispielsweise die Matrix

$$A := \begin{pmatrix} 2 & 3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \quad (3.4)$$

in **R** definieren, so ergibt sich

```
# 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
```

**R** folgt hier *per default* einer sogenannten *column-major-order*, das heißt, die Elemente des Vektors `c(2,1,3,6,0,5)` werden der Reihe nach von oben nach unten in die Spalten der Matrix von links nach rechts überführt. Einen etwas klareren Zusammenhang zwischen dem visuellen Layout des **R**-Codes und der resultierenden Matrix erhält man, indem man den Vektor mithilfe von Zeilenumbrüchen anhand des intendierten Matrixlayouts formatiert und dann die *column-major-order* mithilfe des Arguments `byrow = TRUE` zu einer *row-major-order* umstellt. Es wird dann zunächst die erste Zeile der Matrix von links nach rechts mit den Elementen des Vektors gefüllt wird und dann die zweite Zeile u.s.w. bis alle Elemente des Vektors auf die Matrix verteilt sind.

```
# spaltenweise Definition von A (R-Default)
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

```

```

# 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

```

## 3.2 Operationen

Man kann mit Matrizen rechnen. Dabei sind folgende Matrixoperationen grundlegend:

- Die Addition von Matrizen gleicher Größe, genannt *Matrixaddition*;
- die Subtraktion von Matrizen gleicher Größe, genannt *Matrixsubtraktion*;
- die Multiplikation einer Matrix mit einem Skalar, genannt *Skalarmultiplikation*;
- das Vertauschen der Zeilen und Spalten einer Matrix, genannt *Matrixtransposition*.

Wir führen diese Operationen in der Folge in der Operatorform, also als Funktionen ein. Dies dient insbesondere dazu, bei jeder Operation mit Hilfe ihrer Definitionsmenge zu betonen, von welcher Art die Objekte der jeweiligen Operation sind und mithilfe ihrer Bildmenge zu betonen, von welcher Art das Resultat der jeweiligen Operation ist.

### 3.2.1 Matrixaddition

**Definition 3.2.** Es seien  $A, B \in \mathbb{R}^{n \times m}$ . Dann ist die *Addition* von  $A$  und  $B$  definiert als die Abbildung

$$+ : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto +(A, B) := A + B \quad (3.5)$$

mit

$$\begin{aligned}
 A + B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\
 &:= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}.
 \end{aligned} \quad (3.6)$$

•

Die Definition der Matrixaddition legt insbesondere fest, dass nur Matrizen gleicher Größe addiert werden können und dass die Operation der Matrixaddition elementweise definiert ist, d.h. ein Element der Ergebnismatrix sich aus den korrespondierenden Elementen der Ausgangsmatrizen ergibt.

### Beispiel

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

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix}. \quad (3.7)$$

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

$$\begin{aligned} C = A + B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2+4 & -3+1 & 0+0 \\ 1-4 & 6+2 & 5+0 \end{pmatrix} \\ &= \begin{pmatrix} 6 & -2 & 0 \\ -3 & 8 & 5 \end{pmatrix}. \end{aligned} \quad (3.8)$$

In **R** führt man obige Rechnung wie folgt aus.

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

# Addition
C = A + B
print(C)
```

```
      [,1] [,2] [,3]
[1,]    6  -2    0
[2,]   -3    8    5
```

### 3.2.2 Matrixsubtraktion

Die Subtraktion von Matrizen gleicher Größe ist analog zur Addition definiert.

**Definition 3.3** (Matrixsubtraktion). Es seien  $A, B \in \mathbb{R}^{n \times m}$ . Dann ist die *Subtraktion* von  $A$  und  $B$  definiert als die Abbildung

$$- : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto -(A, B) := A - B \quad (3.9)$$

mit

$$\begin{aligned}
 A - B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\
 &:= \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \cdots & a_{1m} - b_{1m} \\ a_{21} - b_{21} & a_{22} - b_{22} & \cdots & a_{2m} - b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} - b_{n1} & a_{n2} - b_{n2} & \cdots & a_{nm} - b_{nm} \end{pmatrix}.
 \end{aligned} \tag{3.10}$$

•

Wie bei der Matrixaddition legt die Definition der Matrixsubtraktion fest, dass nur Matrizen gleicher Größe voneinander subtrahiert werden können und dass die Subtraktion zweier gleich großer Matrizen elementweise definiert ist.

### Beispiel

Wir können die im Beispiel zur Matrixaddition definierten Matrizen  $A$  und  $B$  auch voneinander subtrahieren,

$$\begin{aligned}
 D = A - B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} - \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} 2 - 4 & -3 - 1 & 0 - 0 \\ 1 + 4 & 6 - 2 & 5 - 0 \end{pmatrix} \\
 &= \begin{pmatrix} -2 & -4 & 0 \\ 5 & 4 & 5 \end{pmatrix}.
 \end{aligned} \tag{3.11}$$

In  $\mathbf{R}$  führt man diese Rechnung wie folgt aus.

```
# Subtraktion
D = A - B
print(D)
```

```
      [,1] [,2] [,3]
[1,]  -2  -4   0
[2,]   5   4   5
```

### 3.2.3 Skalarmultiplikation

Die *Skalarmultiplikation* einer Matrix bezeichnet die Multiplikation einer Matrix mit einem Skalar.

**Definition 3.4** (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

$$\cdot : \mathbb{R} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (c, A) \mapsto \cdot(c, A) := cA \tag{3.12}$$

mit

$$cA = c \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}. \quad (3.13)$$

•

Die Skalarmultiplikation ist mit dieser Definition also ebenfalls elementweise definiert, dahingehend dass jeder Eintrag der Ergebnismatrix sich aus dem Produkt des korrespondierenden Eintrags der Ausgangsmatrix mit dem Skalar ergibt.

### Beispiel

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

$$A := \begin{pmatrix} 3 & 1 & 1 \\ 5 & 2 & 5 \\ 2 & 7 & 1 \\ 3 & 4 & 2 \end{pmatrix}. \quad (3.14)$$

Dann ergibt sich

$$B := cA = -3 \begin{pmatrix} 3 & 1 & 1 \\ 5 & 2 & 5 \\ 2 & 7 & 1 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} -3 \cdot 3 & -3 \cdot 1 & -3 \cdot 1 \\ -3 \cdot 5 & -3 \cdot 2 & -3 \cdot 5 \\ -3 \cdot 2 & -3 \cdot 7 & -3 \cdot 1 \\ -3 \cdot 3 & -3 \cdot 4 & -3 \cdot 2 \end{pmatrix} = \begin{pmatrix} -9 & -3 & -3 \\ -15 & -6 & -15 \\ -6 & -21 & -3 \\ -9 & -12 & -6 \end{pmatrix}. \quad (3.15)$$

In **R** führt man diese Skalarmultiplikation wie folgt aus.

```
# 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
```

Mithilfe der Definition von Matrixaddition und Skalarmultiplikation ist es möglich, einen Vektorraum zu definieren, dessen Elemente die reellen Matrizen sind. Insbesondere legt diese Definition auch die Rechenregeln beim Umgang mit Matrixaddition und Skalarmultiplikation fest.

**Theorem 3.1** (Vektorraum der reellwertigen Matrizen). *Das Tripel  $(\mathbb{R}^{n \times m}, +, \cdot)$  mit der oben definierten Matrixaddition und Skalarmultiplikation ist ein Vektorraum. Insbesondere gelten 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:  $\forall A \exists -A \in \mathbb{R}^{n \times m}$  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:  $(r + s) \cdot A = r \cdot A + s \cdot A$ .*

◦

Wir verzichten auf einen Beweis, der sich mit einigem Notationsaufwand schlicht aus dem elementweisen Charakter von Matrixaddition und Skalarmultiplikation sowie den aus den für die reellen Zahlen bekannten Rechenregeln ergibt. Das im Theorem erwähnte neutrale Element der Skalarmultiplikation ist die *reelle Zahl* 1. Das im Theorem erwähnte neutrale Element der Addition ist die *Nullmatrix*. Das im Theorem erwähnte inverse Element der Addition ist schließlich die Matrix

$$-A := (-a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}. \quad (3.16)$$

Dies erlaubt es, die Matrixsubtraktion als Spezialfall der Matrixaddition zu betrachten.

### 3.2.4 Matrixtransposition

Eine weitere häufig auftretende grundlegende Matrixoperation ist das Vertauschen der Zeilen- und Spaltenanordnung einer Matrix, genannt *Matrixtransposition*.

**Definition 3.5** (Matrixtransposition). Es sei  $A \in \mathbb{R}^{n \times m}$ . Dann ist die *Transposition* von  $A$  definiert als die Abbildung

$$\cdot^T : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{m \times n}, A \mapsto \cdot^T(A) := A^T \quad (3.17)$$

mit

$$A^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}^T := \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}. \quad (3.18)$$

•

Für  $A \in \mathbb{R}^{n \times m}$  gilt damit also immer  $A^T \in \mathbb{R}^{m \times n}$ . Weiterhin gelten folgende Rechenregeln der Matrixtransposition, wie man sich an Beispielen klar macht:

- (1) Für  $A \in \mathbb{R}^{1 \times 1}$  gilt
 
$$A^T = A. \quad (3.19)$$

(2) Es gilt

$$(A^T)^T = A. \quad (3.20)$$

(3) Es gilt

$$(a_{ii})_{1 \leq i \leq \min(n,m)} = (a_{ii})_{1 \leq i \leq \min(n,m)}^T. \quad (3.21)$$

Letztere Eigenschaft der Transposition besagt, dass die Elemente auf der Hauptdiagonalen einer Matrix bei Transposition unberührt bleiben.

### Beispiel

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

$$A := \begin{pmatrix} 2 & 3 & 0 \\ 1 & 6 & 5 \end{pmatrix}. \quad (3.22)$$

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

$$A^T := \begin{pmatrix} 2 & 1 \\ 3 & 6 \\ 0 & 5 \end{pmatrix}. \quad (3.23)$$

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

$$(a_{11}) = (a_{11})^T \text{ und } (a_{22}) = (a_{22})^T. \quad (3.24)$$

In  $\mathbf{R}$  führt man die Transposition einer Matrix wie folgt durch.

```
# 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
```

Schließlich gelten in der Verbindung mit der Matrixaddition, Matrixsubtraktion und der Skalarmultiplikation folgende Rechenregeln, wie man sich an Beispielen klar macht:

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

$$(A + B)^T = A^T + B^T. \tag{3.25}$$

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

$$(A - B)^T = A^T - B^T. \tag{3.26}$$

(3) Für  $c \in \mathbb{R}$  und  $A \in \mathbb{R}^{n \times m}$  gilt

$$(cA)^T = cA^T. \tag{3.27}$$

### 3.3 Matrixmultiplikation

Die Matrixmultiplikation ist die zentrale Operation beim Rechnen mit Matrizen. Sie ist wie folgt definiert.

**Definition 3.6** (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

$$\cdot : \mathbb{R}^{n \times m} \times \mathbb{R}^{m \times k} \rightarrow \mathbb{R}^{n \times k}, (A, B) \mapsto \cdot(A, B) := AB \tag{3.28}$$

mit

$$\begin{aligned}
 AB &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{pmatrix} \\
 &:= \begin{pmatrix} \sum_{i=1}^m a_{1i}b_{i1} & \sum_{i=1}^m a_{1i}b_{i2} & \cdots & \sum_{i=1}^m a_{1i}b_{ik} \\ \sum_{i=1}^m a_{2i}b_{i1} & \sum_{i=1}^m a_{2i}b_{i2} & \cdots & \sum_{i=1}^m a_{2i}b_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ni}b_{i1} & \sum_{i=1}^m a_{ni}b_{i2} & \cdots & \sum_{i=1}^m a_{ni}b_{ik} \end{pmatrix} \\
 &= \left( \sum_{i=1}^m a_{ji}b_{il} \right)_{1 \leq j \leq n, 1 \leq l \leq k}
 \end{aligned} \tag{3.29}$$

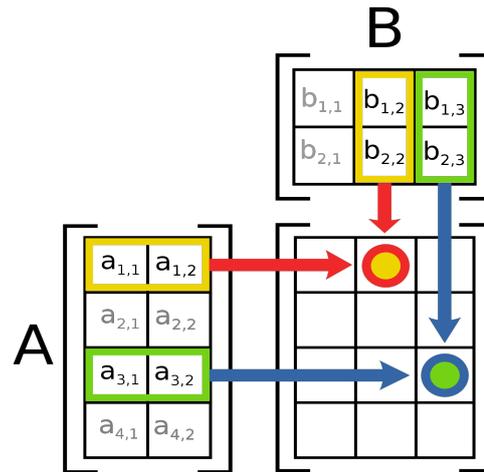
•

Das Matrixprodukt  $AB$  ist also nur dann definiert, wenn  $A$  genau so viele Spalten hat wie  $B$  Zeilen hat. Informell gilt für die beteiligten Matrixgrößen dabei die Merkregel

$$(n \times m)(m \times k) = (n \times k). \tag{3.30}$$

Der Eintrag  $(AB)_{ij}$  in  $AB$  entspricht der Summe der multiplizierten  $i$ ten Zeile von  $A$  und  $j$ ten Spalte von  $B$ . Zum Berechnen von  $(AB)_{ij}$  geht man für  $i = 1, \dots, n$  und  $j = 1, \dots, k$  also in Gedanken wie folgt vor:

- (1) Man legt die Tranposition 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 dann zu jedem Element in der Zeile von  $A$  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  $AB$ .



**Abbildung 3.1** Graphische Darstellung der Matrixmultiplikation (Quelle: [Wikimedia Commons: “Matrix\\_multiplication\\_diagram\\_2.svg”](#); Lizenz: GNU Free, CC-BY-SA 3.0).

Abbildung 3.1 veranschaulicht die Funktionsweise der Matrixmultiplikation.

**Beispiel**

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

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix}. \tag{3.31}$$

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

$$(2 \times 3)(3 \times 2) = (2 \times 2) \tag{3.32}$$

und

$$(3 \times 2)(2 \times 3) = (3 \times 3). \tag{3.33}$$

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

Für  $C$  ergibt sich dann

$$\begin{aligned} C &= AB \\ &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 2 \cdot 4 + (-3) \cdot (-1) + 0 \cdot 1 & 2 \cdot 2 + (-3) \cdot 0 + 0 \cdot 3 \\ 1 \cdot 4 + 6 \cdot (-1) + 5 \cdot 1 & 1 \cdot 2 + 6 \cdot 0 + 5 \cdot 3 \end{pmatrix} \\ &= \begin{pmatrix} 8 + 3 + 0 & 4 + 0 + 0 \\ 4 - 6 + 5 & 2 + 0 + 15 \end{pmatrix} \\ &= \begin{pmatrix} 11 & 4 \\ 3 & 17 \end{pmatrix}. \end{aligned} \tag{3.34}$$

In  $\mathbf{R}$  nutzt man für die Matrixmultiplikation den Operator `%*%`.

```
# 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
```

Für  $D$  ergibt sich weiterhin

$$\begin{aligned}
 D &= BA \\
 &= \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \\
 &= \begin{pmatrix} 4 \cdot 2 + 2 \cdot 1 & 4 \cdot (-3) + 2 \cdot 6 & 4 \cdot 0 + 2 \cdot 5 \\ (-1) \cdot 2 + 0 \cdot 1 & (-1) \cdot (-3) + 0 \cdot 6 & (-1) \cdot 0 + 0 \cdot 5 \\ 1 \cdot 2 + 3 \cdot 1 & 1 \cdot (-3) + 3 \cdot 6 & 1 \cdot 0 + 3 \cdot 5 \end{pmatrix} \\
 &= \begin{pmatrix} 8 + 2 & -12 + 12 & 0 + 5 \\ -2 + 0 & 3 + 0 & 0 + 0 \\ 2 + 3 & -3 + 18 & 0 + 15 \end{pmatrix} \\
 &= \begin{pmatrix} 10 & 0 & 10 \\ -2 & 3 & 0 \\ 5 & 15 & 15 \end{pmatrix}
 \end{aligned} \tag{3.35}$$

In  $\mathbf{R}$  überprüft man diese Rechnung wie folgt.

```
# 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

Ist allerdings eine Matrixmultiplikation aufgrund nicht zueinander passender Matrizengrößen nicht definiert, so lässt sich diese auch in  $\mathbf{R}$  nicht auswerten.

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

Error in t(A) %\*% B: non-conformable arguments

Folgendes Theorem, das wir nicht beweisen wollen, stellt einen Zusammenhang zwischen dem Skalarprodukt zweier Vektoren und der Multiplikation zweier Matrizen her.

**Theorem 3.2** (Matrixmultiplikation und Vektorskalarprodukt). *Es seien  $x, y \in \mathbb{R}^n$ . Dann gilt*

$$\langle x, y \rangle = x^T y. \tag{3.36}$$

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

$$\bar{a}_i := (a_{ji})_{1 \leq j \leq m} \in \mathbb{R}^m \tag{3.37}$$

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

$$\bar{b}_j := (b_{ij})_{1 \leq i \leq m} \in \mathbb{R}^m \tag{3.38}$$

die Spalten von  $B$ , also

$$A^T = (\bar{a}_1 \quad \bar{a}_2 \quad \dots \quad \bar{a}_n) \in \mathbb{R}^{m \times n} \text{ und } B = (\bar{b}_1 \quad \bar{b}_2 \quad \dots \quad \bar{b}_k) \in \mathbb{R}^{m \times k}. \tag{3.39}$$

Dann gilt

$$AB = (\langle \bar{a}_i, \bar{b}_j \rangle)_{1 \leq i \leq n, 1 \leq j \leq k}. \tag{3.40}$$

◦

Der Beweis ergibt sich im Wesentlichen durch die Identifikation von  $\mathbb{R}^n$  und  $\mathbb{R}^{n \times 1}$  und aus der Tatsache, dass nach Definition der Eintrag  $(AB)_{ij}$  im Produkt von  $A \in \mathbb{R}^{n \times m}$  und  $B \in \mathbb{R}^{m \times k}$  dem Vektorskalarprodukt der  $i$ ten Spalte von  $A^T$  und der  $j$ ten Spalte von  $B$  entspricht.

### 3.3.1 Rechenregeln der Matrixmultiplikation

Im Folgenden stellen wir einige grundlegende Rechenregeln der Matrixmultiplikation, insbesondere auch in Kombination mit anderen Matrixoperationen zusammen. Für Beweise der folgenden zwei Theoreme zur Assoziativität und Distributivität, die sich im Wesentlichen mit den entsprechenden Rechenregeln für Summen und Produkte der reellen Zahlen ergeben, verweisen wir auf die weiterführende Literatur.

**Theorem 3.3** (Assoziativität). *Es seien  $A \in \mathbb{R}^{n \times m}$ ,  $B \in \mathbb{R}^{m \times k}$ ,  $C \in \mathbb{R}^{k \times p}$  und  $c \in \mathbb{R}$ . Dann gelten:*

(1) *Die Multiplikation von Matrizen ist assoziativ, es gilt also:*

$$A(BC) = (AB)C. \quad (3.41)$$

(2) *Die Kombination von Matrixmultiplikation und Skalarmultiplikation ist assoziativ:*

$$c(AB) = (cA)B = A(cB). \quad (3.42)$$

◦

Die Assoziativität von Matrizenmultiplikation und Skalarmultiplikation erkennt man leicht bei Betrachtung des  $(j, l)$ ten Elements von  $c(AB)$ ,  $(cA)B$  und  $A(cB)$  als

$$c \left( \sum_{i=1}^m a_{ji} b_{il} \right) = \sum_{i=1}^m (c a_{ji}) b_{il} = \sum_{i=1}^m a_{ji} (c b_{il}). \quad (3.43)$$

**Theorem 3.4** (Distributivität). *Es seien  $A \in \mathbb{R}^{n \times m}$ ,  $B \in \mathbb{R}^{n \times m}$ ,  $C \in \mathbb{R}^{m \times p}$ . Dann gelten*

$$(A + B)C = AC + BC \quad (3.44)$$

und

$$C^T(A + B) = C^T A + C^T B \quad (3.45)$$

◦

Beim Rechnen mit Matrixprodukten lassen sich in Klammern stehende Summen also wie in den reellen Zahlen ausklammern.

**Theorem 3.5** (Nichtkommutativität). *Es seien  $A \in \mathbb{R}^{n \times m}$  und  $B \in \mathbb{R}^{m \times k}$ . Dann gilt im Allgemeinen*

$$AB \neq BA. \quad (3.46)$$

◦

Im Gegensatz zur Multiplikation reeller Zahlen ist die Matrixmultiplikation im Allgemeinen nicht kommutativ.

*Beweis.* Im Fall  $k \neq n$  ist  $BA$  nicht definiert, wir betrachten also nur den Fall  $k = n$ . Wir zeigen durch Angabe eines Gegenbeispiels mit  $A, B \in \mathbb{R}^{2 \times 2}$ , dass im Allgemeinen  $AB = BA$  nicht gilt. Es seien

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ und } B := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{3.47}$$

Dann gilt

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = BA. \tag{3.48}$$

□

**Theorem 3.6** (KMatrixmultiplikation und Transposition). *Es seien  $A \in \mathbb{R}^{m \times n}$  und  $B \in \mathbb{R}^{n \times k}$ . Dann gilt*

$$(AB)^T = B^T A^T. \tag{3.49}$$

◦

*Beweis.* Ein Beweis ergibt sich wie folgt:

$$\begin{aligned} (AB)^T &= \left( \left( \sum_{i=1}^n a_{ji} b_{il} \right)_{1 \leq j \leq m, 1 \leq l \leq k} \right)^T \\ &= \left( \sum_{i=1}^n a_{ij} b_{li} \right)_{1 \leq j \leq m, 1 \leq l \leq k} \\ &= \left( \sum_{i=1}^n b_{li} a_{ij} \right)_{1 \leq l \leq k, 1 \leq j \leq m} \\ &= B^T A^T. \end{aligned} \tag{3.50}$$

□

### 3.4 Matrixinversion

Um den Begriff der inversen Matrix zu motivieren, betrachten wir zunächst das Problem des *Lösens eines linearen Gleichungssystems*. Dazu seien  $A \in \mathbb{R}^{n \times n}$ ,  $x \in \mathbb{R}^n$  und  $b \in \mathbb{R}^n$  und es gelte

$$Ax = b. \tag{3.51}$$

$A$  und  $b$  seien als bekannt vorausgesetzt,  $x$  sei unbekannt. Konkret seien beispielsweise

$$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ und } b := \begin{pmatrix} 5 \\ 11 \end{pmatrix}. \tag{3.52}$$

Dann liegt folgendes lineares Gleichungssystem mit zwei Gleichungen und zwei Unbekannten vor:

$$Ax = b \iff \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \iff \begin{matrix} 1x_1 + 2x_2 & = & 5 \\ 3x_1 + 4x_2 & = & 11 \end{matrix} \tag{3.53}$$

Ziel des Lösens von linearen Gleichungssystemen ist bekanntlich, herauszufinden, für welche  $x$  das Gleichungssystem erfüllt ist. Um in diesem Kontext den Begriff der inversen Matrix von  $A$  einzuführen, vereinfachen wir die Situation weiter. Wir nehmen an, dass

$A = a$  eine  $1 \times 1$  Matrix, also ein Skalar, sei und ebenso  $x$  und  $b$ , dass wir also für  $a, x, b \in \mathbb{R}$  folgende Gleichung betrachten:

$$ax = b \quad (3.54)$$

Um diese Gleichung nach  $x$  aufzulösen würde man natürlich beide Seiten der Gleichung mit dem *multiplikativen Inversen* von  $a$  multiplizieren, wobei das *multiplikative Inverse* von  $a$  den Wert bezeichnet, der mit  $a$  multipliziert 1 ergibt. Dieser ist bekanntlich durch

$$a^{-1} = \frac{1}{a} \quad (3.55)$$

gegeben. Dann würde nämlich gelten

$$ax = b \Leftrightarrow a^{-1}ax = a^{-1}b \Leftrightarrow 1 \cdot x = a^{-1}b \Leftrightarrow x = \frac{b}{a} \quad (3.56)$$

Konkret etwa

$$2x = 6 \Leftrightarrow 2^{-1}2x = 2^{-1}6 \Leftrightarrow \frac{1}{2}2x = \frac{1}{2}6 \Leftrightarrow x = 3. \quad (3.57)$$

Analog zu dem Fall, dass die Matrizen in  $Ax = b$  allesamt Skalare sind, möchte man im Fall eines linearen Gleichungssystems beide Seiten der Gleichung mit dem *multiplikativen Inversen*  $A^{-1}$  von  $A$  multiplizieren können, sodass eine Gleichung der Form

$$A^{-1}A = "1". \quad (3.58)$$

resultiert. Dann hätte man nämlich

$$Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b. \quad (3.59)$$

Diese intuitive Idee des multiplikativen Inversen einer Matrix  $A$  wird im Folgenden unter dem Begriff der *inversen Matrix* formalisiert. Dazu benötigen wir zunächst den Begriff der *Einheitsmatrix*.

**Definition 3.7** (Einheitsmatrix). Die Matrix

$$I_n := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \quad (3.60)$$

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

•

In  $\mathbf{R}$  wird  $I_n$  mit dem Befehl `diag(n)` erzeugt. Die Einheitsmatrix ist für die Matrixmultiplikation das Analog zur 1 bei der Multiplikation reeller Zahlen. Das ist die Aussage folgenden Theorems.

**Theorem 3.7** (Neutrales Element der Matrixmultiplikation).  $I_n$  ist das neutrale Element der Matrixmultiplikation, das heißt es gilt für  $A \in \mathbb{R}^{n \times m}$ , dass

$$I_n A = A \text{ und } A I_m = A. \quad (3.61)$$

◦

*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  $1 \leq j \leq m$

$$b_{ij} = 0 \cdot a_{1j} + 0 \cdot a_{2j} + \dots + 0 \cdot a_{i-1,j} + 1 \cdot a_{ij} + 0 \cdot a_{i+1,j} + \dots + 0 \cdot a_{nj} = a_{ij}. \quad (3.62)$$

Analog zeigt man dies für  $A I_m$ . □

Mit dem Begriff der Einheitsmatrix können wir jetzt die Begriffe der inversen Matrix und der invertierbaren Matrix definieren.

**Definition 3.8** (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, sodass

$$A^{-1} A = A A^{-1} = I_n \quad (3.63)$$

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

Man beachte, dass sich die Begriffe der inversen Matrix und der Invertierbarkeit *nur* auf quadratische Matrizen beziehen. Insbesondere können quadratische Matrizen invertierbar sein, müssen es aber nicht sein. (Lineare Gleichungssysteme können eindeutige Lösungen haben, müssen es aber nicht.) Es gibt also quadratische Matrizen, die invertierbar sind, und solche, die es nicht sind. Nicht-invertierbare Matrizen nennt man auch *singuläre* Matrizen, invertierbare Matrizen manchmal auch *nicht-singuläre* Matrizen. Schließlich beachte man, dass Definition 3.8 lediglich aussagt, was eine inverse Matrix ist, aber nicht wie man sie berechnet.

### Beispiel für eine invertierbare Matrix

Die Matrix

$$A := \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \quad (3.64)$$

ist invertierbar und ihre inverse Matrix ist gegeben durch

$$A^{-1} = \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix}, \quad (3.65)$$

denn

$$\begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix} \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix}, \quad (3.66)$$

wovon man sich durch Nachrechnen überzeugt.

### Beispiel für eine nicht-invertierbare Matrix

Die Matrix

$$B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (3.67)$$

ist nicht invertierbar, denn wäre  $B$  invertierbar, dann gäbe es eine Matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (3.68)$$

sodass

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.69)$$

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

### 3.4.1 Zum Berechnen inverser Matrizen

$2 \times 2$  bis etwa  $5 \times 5$  Matrizen kann man prinzipiell per Hand invertieren, dazu stellt die Lineare Algebra verschiedene Verfahren bereit. Wir wollen hier auf eine Einführung in die Matrizeninvertierung per Hand verzichten, da in der Anwendung Matrizen standardmäßig numerisch invertiert werden. Die numerische Matrixinversion ist dann auch ein großes Feld der Forschung zur Numerischen Mathematik, die eine Vielzahl von Algorithmen zu diesem Zweck bereitstellt. In  $\mathbf{R}$  werden Matrizen *per default* mit der Funktion `solve()`, in Anlehnung an das Lösen linearer Gleichungssysteme, invertiert. Für das obige Beispiel einer invertierbaren Matrix ergibt sich dabei folgender  $\mathbf{R}$ -Code.

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

# Berechnen von A^{-1}
print(solve(A))
```

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

```
# Überprüfen der Eigenschaften einer inversen Matrix
print(solve(A) %% A)
```

```
      [,1] [,2]
[1,]    1    0
[2,]    0    1
```

```
# Bei der umgekehrten Berechnung ergeben sich kleine Rundungsfehler
print(A %% solve(A))
```

```

      [,1]      [,2]
[1,] 1 -5.551115e-17
[2,] 0 1.000000e+00
    
```

Nicht-invertierbare Matrizen sind dabei natürlich auch numerisch nicht-invertierbar, wie folgende Fehlermeldung in **R** bezüglich obigen Beispiels einer nicht-invertierbaren Matrix demonstriert.

```

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

Error in solve.default(B): Lapack routine dgesv: system is exactly singular: U[2,2] = 0

### 3.5 Determinanten

Die Determinante ist eine vielseitig einsetzbare Maßzahl einer quadratischen Matrix. Für das Verständnis der *Eigenanalyse* und der *Matrixzerlegung* ist der Begriff der Determinante im Kontext des *charakteristischen Polynoms* grundlegend.

Allgemein ist eine Determinante eine nichtlineare Abbildung der Form

$$|\cdot| : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}, A \mapsto |A|, \tag{3.70}$$

das heißt, eine Determinante ordnet einer quadratischen Matrix  $A$  die reelle Zahl  $|A|$  zu. Die Zahl  $|A|$  wird dabei rekursiv anhand folgender Definition bestimmt.

**Definition 3.9** (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

$$\begin{aligned} |A| &:= a_{11} && \text{für } n = 1 \\ |A| &:= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| && \text{für } n > 1 \end{aligned} \tag{3.71}$$

die *Determinante* von  $A$ .



Die Definition führt die Bestimmung der Determinante einer quadratischen Matrix also sukzessive durch Streichen von Zeilen und Spalten auf die Determinante einer  $1 \times 1$  Matrix zurück, die durch ihr einziges Element gegeben ist. Für

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \tag{3.72}$$

ergeben sich dabei zum Beispiel folgende Matrizen der Form  $A_{ij} \in \mathbb{R}^{2 \times 2}$ :

$$A_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, A_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, A_{21} = \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}, A_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}. \tag{3.73}$$

Für die Berechnung der Determinanten von zwei- und dreidimensionalen quadratischen Matrizen gibt es direkte, nicht-rekursive Rechenregeln, die in folgendem Theorem festgehalten sind.

**Theorem 3.8** (Determinanten von zwei- und dreidimensionalen Matrizen).

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

$$|A| = a_{11}a_{22} - a_{12}a_{21}. \quad (3.74)$$

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

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \quad (3.75)$$

◦

*Beweis.*

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

$$\begin{aligned} |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\ &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| \\ &= a_{11}|(a_{22})| - a_{12}|(a_{21})| \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned} \quad (3.76)$$

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

$$\begin{aligned} |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\ &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| + a_{13}(-1)^{1+3}|A_{13}| \\ &= a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| \\ &= a_{11} \left| \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \right| - a_{12} \left| \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} \right| + a_{13} \left| \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \right| \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \end{aligned} \quad (3.77)$$

□

Für die Bestimmung der Determinanten von  $2 \times 2$  und  $3 \times 3$  Matrizen gilt somit die sogenannte *Sarrusche Merkregel*:

“Summe der Produkte auf den Diagonalen minus Summe der Produkte auf den Gegendiagonalen.”

Bei  $3 \times 3$  Matrizen bezieht sich die Merkregel auf ein Schema, das man erhält, wenn man in Gedanken die ersten zwei Spalten der Matrix rechts der Matrix noch einmal aufschreibt:

$$\left( \begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \right). \quad (3.78)$$

### Beispiele für Determinanten von $2 \times 2$ und $3 \times 3$ Matrizen

Es seien

$$A := \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ und } C := \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad (3.79)$$

Dann ergeben sich

$$|A| = 2 \cdot 4 - 1 \cdot 3 = 8 - 3 = 5 \quad (3.80)$$

und

$$|B| = 1 \cdot 0 - 0 \cdot 0 = 0 - 0 = 0 \quad (3.81)$$

und

$$|C| = 2 \cdot 1 \cdot 3 + 0 \cdot 0 \cdot 0 + 0 \cdot 0 \cdot 0 - 0 \cdot 0 \cdot 3 - 0 \cdot 0 \cdot 0 - 0 \cdot 1 \cdot 0 = 2 \cdot 1 \cdot 3 = 6. \quad (3.82)$$

In  $\mathbf{R}$  rechnet man dies mithilfe der Funktion `det()` wie folgt nach.

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

```
[1] 5
```

```
# Matrixdefinition und Determinantenberechnung
B = matrix(c(1,0,
            0,0),
          nrow = 2,
          byrow = TRUE)
det(B)
```

```
[1] 0
```

```
# Matrixdefinition und Determinantenberechnung
C = matrix(c(2,0,0,
            0,1,0,
            0,0,3),
          nrow = 3,
          byrow = TRUE)
det(C)
```

```
[1] 6
```

Für Determinanten bestehen zahlreiche Rechenregeln im Zusammenspiel mit Matrixmultiplikation und Matrixinversion. Ohne Beweis stellen wir diese in folgendem Theorem zusammen.

**Theorem 3.9** (Rechenregeln für Determinanten).

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

$$|AB| = |A||B|. \tag{3.83}$$

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

$$|A| = |A^T|. \tag{3.84}$$

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

$$|A^{-1}| = \frac{1}{|A|}. \tag{3.85}$$

(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

$$|A| = \prod_{i=1}^n a_{ii}. \tag{3.86}$$

◦

Dreiecksmatrizen sind hierbei Matrizen, bei denen alle Einträge oberhalb der Diagonale oder alle Einträge unterhalb der Diagonale Null sind oder beides der Fall ist. Einheitsmatrizen können demzufolge als Spezialfälle von Dreiecksmatrizen gesehen werden und ihre Determinanten sind immer Eins:  $|I_n| = \prod_{i=1}^n 1 = 1$ .

Folgendes sehr tiefgehendes Theorem, welches wir nicht vollständig beweisen wollen, gibt eine Möglichkeit an, anhand der Determinante einer quadratischen Matrix zu bestimmen, ob sie invertierbar ist.

**Theorem 3.10** (Invertierbarkeit und Determinante).  $A \in \mathbb{R}^{n \times n}$  ist dann und nur dann invertierbar, wenn gilt, dass  $|A| \neq 0$ . Es gilt also

$$A \text{ ist invertierbar} \Leftrightarrow |A| \neq 0 \quad \text{und} \quad A \text{ ist nicht invertierbar} \Leftrightarrow |A| = 0. \tag{3.87}$$

◦

*Beweis.* Wir deuten einen Beweis lediglich an und zeigen, dass aus der Invertierbarkeit von  $A$  folgt, dass  $|A|$  nicht gleich 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

$$|AB| = |A||B| = |I_n| = 1. \tag{3.88}$$

Also kann  $|A| = 0$  nicht gelten, denn sonst wäre  $0 = 1$ . □

### 3.5.1 Visuelle Intuition

Der abstrakte Begriff der Determinante einer quadratischen Matrix kann mithilfe des Vektorraumbegriffs etwas veranschaulicht werden. Dazu seien  $a_1, \dots, a_n \in \mathbb{R}^n$  die Spalten von  $A \in \mathbb{R}^{n \times n}$ . Dann entspricht  $|A|$  dem signierten Volumen des von  $a_1, \dots, a_n \in \mathbb{R}^n$  aufgespannten Parallelotops. Um dies visuell zu veranschaulichen betrachten wir die Matrizen

$$A_1 = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \tag{3.89}$$

mit den jeweiligen Determinanten

$$|A_1| = 3 \cdot 2 - 1 \cdot 1 = 5, \quad |A_2| = 2 \cdot 2 - 0 \cdot 0 = 4, \quad |A_3| = 2 \cdot 2 - 2 \cdot 2 = 0. \tag{3.90}$$

Abbildung 3.2 visualisiert die entsprechende Intuition.

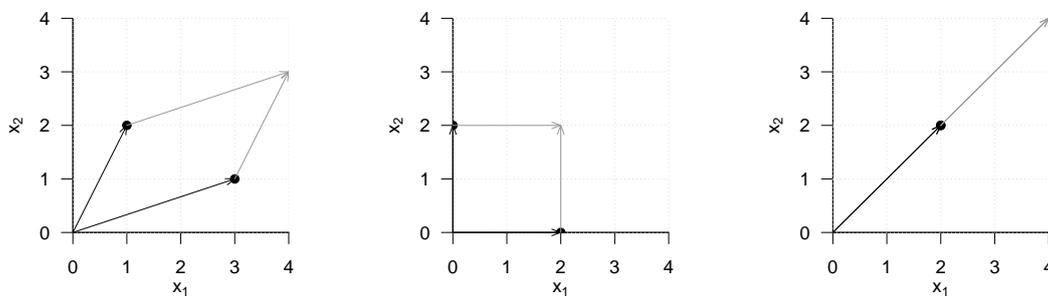


Abbildung 3.2 Determinanten als Flächeninhalte von Parallelogrammen im Fall  $n = 2$ .

## 3.6 Spezielle Matrizen

In diesem Abschnitt stellen wir einige häufig auftretende Typen von Matrizen und ihre Eigenschaften zusammen. Zum Beweis der allermeisten Eigenschaften verweisen wir dabei auf die weiterführende Literatur.

### 3.6.1 Einheitsmatrizen

Die Einheitsmatrix haben wir bereits kennengelernt. Wir fassen sie hier mit den Einheitsvektoren noch einmal in einer gemeinsamen Definition zusammen.

**Definition 3.10** (Einheitsmatrix und Einheitsvektoren).

Wir bezeichnen die *Einheitsmatrix* mit

$$I_n := (i_{jk})_{1 \leq j \leq n, 1 \leq k \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ f\u00fcr } j = k \text{ und } i_{jk} = 0 \text{ f\u00fcr } j \neq k. \tag{3.91}$$

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

$$e_i := (e_{i_j})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{i_j} = 1 \text{ f\u00fcr } i = j \text{ und } e_{i_j} = 0 \text{ f\u00fcr } i \neq j. \tag{3.92}$$

•

Die Einheitsmatrix  $I_n$  besteht nur aus Nullen und Diagonalelementen gleich Eins, die Einheitsvektoren bestehen nur aus Nullen und einer Eins in der jeweils indizierten Komponente. Es gilt

$$I_n = (e_1 \ \dots \ e_n) \in \mathbb{R}^{n \times n}. \quad (3.93)$$

Für  $n = 3$  gilt also zum Beispiel

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ und } e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.94)$$

Weiterhin gelten für die Einheitsvektoren für  $1 \leq i, j \leq n$

$$e_i^T e_j = 0 \text{ für } i \neq j, e_i^T e_i = 1 \text{ und } e_i^T v = v^T e_i = v_i \text{ für } v \in \mathbb{R}^n. \quad (3.95)$$

### 3.6.2 Einsmatrizen und Nullmatrizen

**Definition 3.11** (Nullmatrizen, Nullvektoren, Einsmatrizen, Einsvektoren).

Wir bezeichnen *Nullmatrizen* und *Nullvektoren* mit

$$0_{nm} := (0)_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n. \quad (3.96)$$

Wir bezeichnen *Einsmatrizen* und *Einsvektoren* mit

$$1_{nm} := (1)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 1_n := (1)_{1 \leq i \leq n} \in \mathbb{R}^n. \quad (3.97)$$

•

$0_{nm}$  und  $0_n$  bestehen also nur aus Nullen und  $1_{nm}$  und  $1_n$  bestehen nur aus Einsen. Es gilt also beispielsweise

$$0_{32} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, 0_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, 1_{32} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \text{ und } 1_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad (3.98)$$

Weiterhin gelten zum Beispiel

$$0_n 0_n^T = 0_{nn} \text{ und } 1_n 1_n^T = 1_{nn}, \quad (3.99)$$

wovon man sich durch Nachrechnen überzeugt.

### 3.6.3 Diagonalmatrizen

**Definition 3.12** (Diagonalmatrix). Eine Matrix  $D \in \mathbb{R}^{n \times m}$  heißt *Diagonalmatrix*, wenn  $d_{ij} = 0$  für  $1 \leq i \leq n, 1 \leq j \leq m$  mit  $i \neq j$ .

•

Eine Diagonalmatrix ist also eine Matrix, für die alle Einträge abseits der Hauptdiagonalen Null sind. Über die Einträge auf der Hauptdiagonalen ist damit nichts gesagt. Eine quadratische Diagonalmatrix  $D \in \mathbb{R}^{n \times n}$  mit den Diagonalelementen  $d_1, \dots, d_n \in \mathbb{R}$  schreibt man auch als

$$D = \text{diag}(d_1, \dots, d_n). \quad (3.100)$$

Zum Beispiel gelten

$$D := \text{diag}(1, 2, 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad (3.101)$$

und für  $\sigma^2 \in \mathbb{R}$

$$\Sigma = \text{diag}(\sigma^2, \sigma^2, \sigma^2) = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_3. \quad (3.102)$$

In folgendem Theorem stellen wir einige wichtige Eigenschaften von quadratischen Diagonalmatrizen zusammen.

**Theorem 3.11** (Eigenschaften quadratischer Diagonalmatrizen).

(Determinante.)  $D := \text{diag}(d_1, \dots, d_n) \in \mathbb{R}^{n \times n}$  sei eine quadratische Diagonalmatrix. Dann gilt

$$|D| = \prod_{i=1}^n d_i. \quad (3.103)$$

◦

### 3.6.4 Symmetrische Matrizen

**Definition 3.13.** Eine Matrix  $S \in \mathbb{R}^{n \times n}$  heißt *symmetrisch*, wenn  $S^T = S$ .

•

Symmetrische Matrizen sind also quadratische Matrizen, die bei Transposition unverändert bleiben. Ein Beispiel für eine symmetrische Matrix ist

$$S := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}. \quad (3.104)$$

In folgendem Theorem stellen wir einige wichtige Eigenschaften symmetrischer Matrizen zusammen.

**Theorem 3.12** (Eigenschaften symmetrischer Matrizen).

(Summation.)  $A \in \mathbb{R}^{n \times n}$  und  $B \in \mathbb{R}^{n \times n}$  seien symmetrische Matrizen. Dann gilt

$$A + B = (A + B)^T. \quad (3.105)$$

(Inverse.)  $S$  sei eine invertierbare symmetrische Matrix und  $S^{-1}$  ihre Inverse. Dann ist auch  $S^{-1}$  eine symmetrische Matrix, das heißt es gilt

$$(S^{-1})^T = S^{-1}. \quad (3.106)$$

◦

### 3.6.5 Orthogonale Matrizen

**Definition 3.14.** Eine Matrix  $Q \in \mathbb{R}^{n \times n}$  heißt *orthogonal*, wenn  $Q^T Q = I_n$ .

•

Die Spalten einer orthogonalen Matrix sind also paarweise orthogonal, es gilt für

$$Q = (q_1 \ \cdots \ q_n) \text{ mit } q_i \in \mathbb{R}^n \text{ für } 1 \leq i \leq n, \quad (3.107)$$

dass

$$q_i^T q_j = 0 \text{ für } i \neq j \text{ und } q_i^T q_j = 1 \text{ für } i = j \text{ mit } 1 \leq i, j \leq n. \quad (3.108)$$

In folgendem Theorem stellen wir einige wichtige Eigenschaften symmetrischer Matrizen zusammen.

**Theorem 3.13** (Eigenschaften orthogonaler Matrizen).  $Q \in \mathbb{R}^{n \times n}$  sei eine orthogonale Matrix. Dann gelten folgende Eigenschaften von  $Q$ .

(Inverse.) Die Inverse von  $Q$  ist  $Q^T$ , es gilt

$$Q^{-1} = Q^T. \quad (3.109)$$

(Transposition.) Die Zeilen von  $Q$  sind orthonormal, es gilt

$$Q Q^T = I_n \quad (3.110)$$

◦

*Beweis.*

(Inverse.) Unter der Annahme, dass  $Q^{-1}$  existiert, gilt

$$Q^T Q = I_n \Leftrightarrow Q^T Q Q^{-1} = I_n Q^{-1} \Leftrightarrow Q^{-1} = Q^T. \quad (3.111)$$

(Transposition.) Es gilt

$$Q^T Q = I_n \Leftrightarrow Q Q^T Q = Q I_n \Leftrightarrow Q Q^T Q Q^T = Q Q^T \Leftrightarrow Q Q^T = I_n. \quad (3.112)$$

□

### 3.6.6 Positiv-definite Matrizen

**Definition 3.15.** 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.

•

Positiv-definite Matrizen sind für die probabilistische Modellbildung unter Verwendung multivariater Normalverteilungen zentral.

In folgendem Theorem stellen wir einige wichtige Eigenschaften positiv-definiter Matrizen zusammen.

**Theorem 3.14** (Eigenschaften positiv-definiter Matrizen).

(Inverse.)  $C \in \mathbb{R}^{n \times n}$  sei eine positiv-definite Matrix. Dann gilt, dass  $C^{-1}$  existiert und ebenfalls positiv-definit ist.

◦

## 3.7 Literaturhinweise

S. Searle (1982) gibt eine umfassende Einführung in die Matrixtheorie vor dem Hintergrund der probabilistischen Datenanalyse, Strang (2009) gibt eine umfassende Einführung in die Matrixtheorie im Kontext der linearen Algebra. In ihrer modernen Inkarnation tauchen Matrizen als algebraische Objekte wohl zunächst in den Arbeiten von Arthur Cayley (1821-1895) auf, siehe zum Beispiel Cayley (1858).

## 3.8 Selbstkontrollfragen

1. Geben Sie die Definition einer Matrix wieder.
2. Nennen Sie sechs Matrixoperationen.
3. Geben Sie die 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

$$A := \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad B := \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix} \quad \text{und} \quad c := 2 \quad (3.113)$$

Berechnen Sie

$$D := c(A - B^T) \quad \text{und} \quad E := (cA)^T + B \quad (3.114)$$

per Hand.

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:

$$ABC, \quad ABC^T, \quad A^T C B^T, \quad BAC. \quad (3.115)$$

9. Es seien

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 2 & 0 \end{pmatrix}, B := \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 0 & 0 \end{pmatrix} \text{ und } C := \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}. \quad (3.116)$$

Berechnen Sie die Matrixprodukte

$$AB, \quad B^T A^T, \quad (B^T A^T)^T, \quad AC \quad (3.117)$$

per Hand.

10. Definieren Sie die Begriffe der invertierbaren Matrix und der inversen Matrix.
11. Geben Sie die Definition von Einheitsmatrix und Einheitsvektoren wieder.
12. Geben Sie die Definition von Einsmatrizen und Nullmatrizen wieder.
13. Geben Sie die Definition einer symmetrischen Matrix wieder.
14. Geben Sie die Definition einer Diagonalmatrix wieder.
15. Geben Sie die Definition einer positiv-definiten Matrix wieder.

## 4 Normalverteilungen

Die multivariate Normalverteilung ist die multivariate Generalisierung der univariaten Normalverteilung. Die Motivation für die verbreiteten Normalverteilungsannahmen in der probabilistischen Modellierung, insbesondere in linearen Modellen, liegt bekanntermaßen im Zentralen Grenzwertsatz: In probabilistischen Modellen repräsentieren probabilistische Terme die Summation sehr vieler Zufallsvorgänge, die durch die deterministischen Bestandteile des jeweiligen Modells, also eine mechanistische wissenschaftliche Theorie, nicht erklärt werden. Nach dem Zentralen Grenzwertsatz ist die Summe dieser nicht erklärten Zufallsvorgänge dann aber gerade normalverteilt.

Über diesen grundlegenden Aspekt hinaus hat die Normalverteilung viele günstige mathematische Eigenschaften, die ihren Einsatz in vielen Bereichen der probabilistischen Modellierung ermöglichen. Anwendungen multivariater Normalverteilungen finden sich damit im Kontext des Allgemeinen Linearen Modells, dessen Generalisierungen zu Hierarchischen Linearen Modellen oder Multivariaten Allgemeinen Linearen Modellen, der Bayesianischen Inferenz bei Normalverteilungsannahmen und nicht zuletzt der Theorie probabilistischer Filter, wie zum Beispiel des Kalman-Bucy-Filters.

### 4.1 Konstruktion

In diesem Abschnitt wollen wir zeigen, wie ein bivariat normalverteilter Zufallsvektor durch Transformation und Konkatenation zweier univariat normalverteilter Zufallsvariablen konstruiert werden kann. Dazu erinnern wir zunächst an den Begriff der normalverteilten Zufallsvariable.

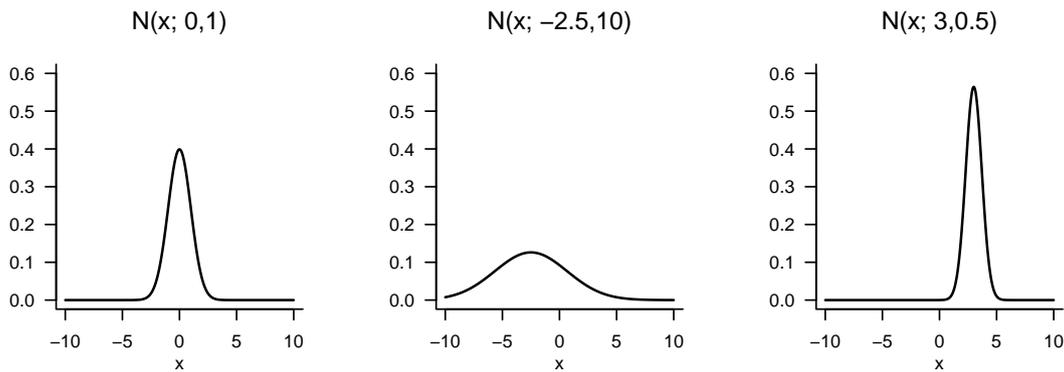
**Definition 4.1** (Normalverteilte Zufallsvariable).  $\xi$  sei eine Zufallsvariable mit Ergebnisraum  $\mathbb{R}$  und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (4.1)$$

Dann sagen wir, dass  $\xi$  einer *Normalverteilung* (oder *Gauß-Verteilung*) mit *Erwartungswertparameter*  $\mu \in \mathbb{R}$  und *Varianzparameter*  $\sigma^2 > 0$  unterliegt und nennen  $\xi$  eine *normalverteilte Zufallsvariable*. Wir kürzen dies mit  $\xi \sim N(\mu, \sigma^2)$  ab. Die WDF einer normalverteilten Zufallsvariable bezeichnen wir mit

$$N(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (4.2)$$

•



**Abbildung 4.1** Wahrscheinlichkeitsdichtefunktionen univariater Normalverteilungen.

Visuell entspricht der Parameter  $\mu$  einer normalverteilten Zufallsvariable dem Wert höchster Wahrscheinlichkeitsdichte und der Parameter  $\sigma^2$  spezifiziert die Breite der WDF (siehe Abbildung 4.1). Weiterhin gelten für den Erwartungswert und die Varianz einer normalverteilten Zufallsvariable bekanntlich

$$\mathbb{E}(\xi) = \mu \text{ und } \mathbb{V}(\xi) = \sigma^2. \quad (4.3)$$

Eine normalverteilte Zufallsvariable der Form  $\xi \sim N(0,1)$  schließlich heißt auch *standardnormalverteilt*. Folgendes Theorem zeigt, wie zwei unabhängige, univariat standardnormalverteilte Zufallsvariablen kombiniert werden können, um einen bivariat verteilten Zufallsvektor zu konstruieren. Die Verteilung eines ebensolchen Zufallsvektors wird dann als *bivariate Normalverteilung* bezeichnet.

**Theorem 4.1** (Konstruktion bivariater Normalverteilungen).  $\zeta_1 \sim N(0,1)$  und  $\zeta_2 \sim N(0,1)$  seien zwei unabhängige standardnormalverteilte Zufallsvariablen. Weiterhin seien  $\mu_1, \mu_2 \in \mathbb{R}$ ,  $\sigma_1, \sigma_2 > 0$  und  $\rho \in ]-1, 1[$ . Schließlich seien

$$\begin{aligned} \xi_1 &:= \sigma_1 \zeta_1 + \mu_1 \\ \xi_2 &:= \sigma_2 (\rho \zeta_1 + (1 - \rho^2)^{1/2} \zeta_2) + \mu_2. \end{aligned} \quad (4.4)$$

Dann hat die WDF des Zufallsvektors  $\xi := (\xi_1, \xi_2)^T$ , also der gemeinsamen Verteilung von  $\xi_1$  und  $\xi_2$ , die Form

$$p : \mathbb{R}^2 \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad (4.5)$$

wobei  $n := 2$  und  $\mu \in \mathbb{R}^2$  sowie  $\Sigma \in \mathbb{R}^{2 \times 2}$  durch

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ und } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_2 \sigma_1 & \sigma_2^2 \end{pmatrix} \quad (4.6)$$

gegeben sind.

◦

Für einen Beweis des Theorems verweisen wir auf DeGroot und Schervish (2012).

### Beispiel

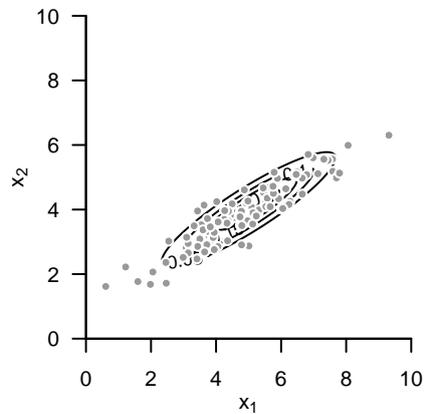


Abbildung 4.2 Konstruktion bivariater Normalverteilungen.

Folgender **R**-Code zeichnet das obige Theorem anhand konkreter Beispielwerte für  $\mu_1, \mu_2, \sigma_1, \sigma_2$  und  $\rho$  nach und gibt die Parameter  $\mu$  und  $\Sigma$  der resultierenden bivariaten Normalverteilung aus.

```
# Parameterdefinitionen
mu_1 = 5.0 # \mu_1
mu_2 = 4.0 # \mu_2
sig_1 = 1.5 # \sigma_1
sig_2 = 1.0 # \sigma_2
rho = 0.9 # \rho

# Realisierungen der standardnormalverteilten ZVen
n = 100 # Anzahl Realisierungen
zeta_1 = rnorm(n) # \zeta_1 \sim N(0,1)
zeta_2 = rnorm(n) # \zeta_2 \sim N(0,1)

# Evaluation der Realisierungen von \xi_1 und \xi_2
xi_1 = sig_1*zeta_1 + mu_1 # Realisierungen von \xi_1
xi_2 = sig_2*(rho*zeta_1 + sqrt(1-rho^2)*zeta_2) + mu_2 # Realisierungen von \xi_2

# Parameter der gemeinsamen Verteilung von \xi_1 und \xi_2
mu = matrix(c(mu_1, # \mu \in \mathbb{R}^2
              mu_2),
            nrow = 2, byrow = TRUE)
Sigma = matrix(c(sig_1^2, rho*sig_1*sig_2, # \Sigma \in \mathbb{R}^{2 x 2}
                 rho*sig_1*sig_2, sig_2^2),
              nrow = 2, byrow = TRUE)

print(mu)
```

```
      [,1]
[1,]    5
[2,]    4
```

```
print(Sigma)
```

```

      [,1] [,2]
[1,] 2.25 1.35
[2,] 1.35 1.00

```

Die durch obigen **R**-Code generierten Realisierungen von  $\xi = (\xi_1, \xi_2)^T$  sowie die Isokonturen der durch Theorem 4.1 vorhergesagten WDF sind in Abbildung 4.2 dargestellt.

## 4.2 Definition

Wir wollen die multivariate Normalverteilung nun formal einführen und erste Eigenschaften angeben. Wir nutzen dazu folgende Definition.

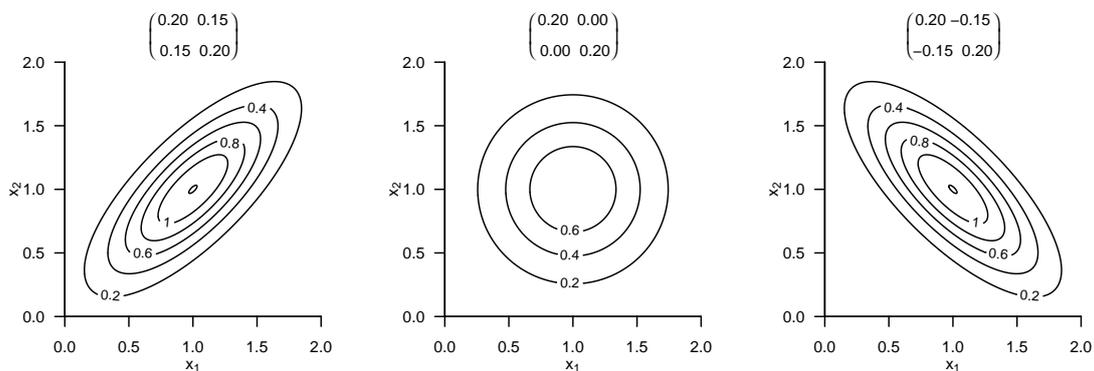
**Definition 4.2.**  $\xi$  sei ein  $n$ -dimensionaler Zufallsvektor mit Ergebnisraum  $\mathbb{R}^n$  und WDF

$$p : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right). \quad (4.7)$$

Dann sagen wir, dass  $\xi$  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  $\xi$  einen *(multivariat) normalverteilten Zufallsvektor*. Wir kürzen dies mit  $\xi \sim N(\mu, \Sigma)$  ab. Die WDF eines multivariat normalverteilten Zufallsvektors bezeichnen wir mit

$$N(x; \mu, \Sigma) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right). \quad (4.8)$$

•

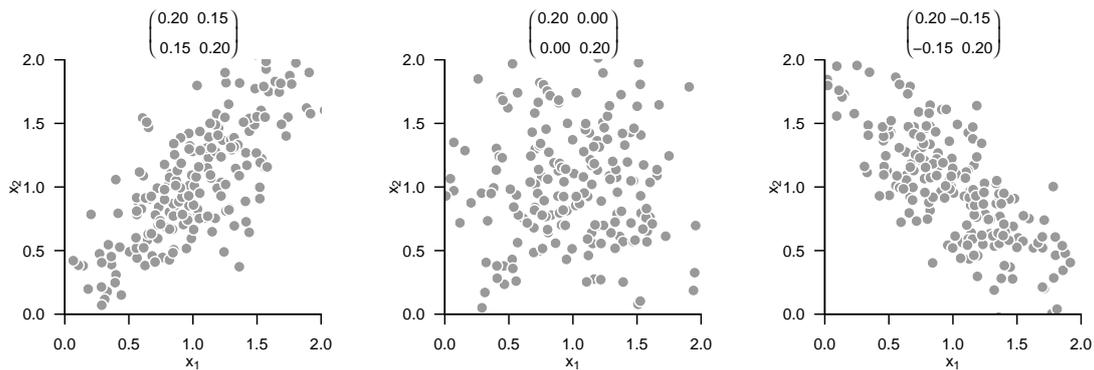


**Abbildung 4.3** Wahrscheinlichkeitsdichtefunktionen bivariater Normalverteilungen.

Abbildung 4.3 A,B,C zeigen die Isokonturen der Wahrscheinlichkeitsdichtefunktionen bivariat normalverteilter Zufallsvektoren für  $\mu = (1, 1)^T$  und

$$\Sigma_A := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix}, \quad \Sigma_B := \begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.20 \end{pmatrix}, \quad \Sigma_C := \begin{pmatrix} 0.20 & -0.15 \\ -0.15 & 0.20 \end{pmatrix}. \quad (4.9)$$

Abbildung 4.4 A,B,C zeigen jeweils 200 Realisierungen der entsprechenden bivariat normalverteilten Zufallsvektoren.



**Abbildung 4.4** Realisierungen bivariat normalverteilter Zufallsvektoren.

Ohne Beweis halten wir fest, dass, wie im Fall einer univariat normalverteilten Zufallsvariable, der Erwartungswert und die Kovarianzmatrix eines multivariat normalverteilten Zufallsvektors durch die entsprechenden Parameter gegeben sind.

**Theorem 4.2** (Erwartungswert und Kovarianzmatrix normalverteilter Zufallsvektoren).  $\xi \sim N(\mu, \Sigma)$  sei ein multivariat normalverteilter Zufallsvektor mit Erwartungswertparameter  $\mu \in \mathbb{R}^n$  und Kovarianzmatrixparameter  $\Sigma \in \mathbb{R}^{n \times n}$  p.d.. Dann gelten

$$\mathbb{E}(\xi) = \mu \text{ und } \mathbb{C}(\xi) = \Sigma. \quad (4.10)$$

◦

Wie im Falle der univariat normalverteilten Zufallsvariable entspricht der Parameter  $\mu \in \mathbb{R}^n$  dem Wert höchster Wahrscheinlichkeitsdichte der multivariaten Normalverteilung. Analog zum Varianzparameter der univariat normalverteilten Zufallsvariable spezifizieren die Diagonalelemente von  $\Sigma \in \mathbb{R}^{n \times n}$  p.d. die Breite der WDF bezüglich der Zufallsvektorkomponenten  $\xi_1, \dots, \xi_n$ . Allgemein spezifiziert im Falle des multivariat normalverteilten Zufallsvektors das  $i, j$ te Element von  $\Sigma \in \mathbb{R}^{n \times n}$  p.d. hier nun die Kovarianz der Zufallsvektorkomponenten  $\xi_i$  und  $\xi_j$ .

### 4.3 Transformationen

In diesem Abschnitt stellen wir einige Resultate zu den Verteilungen transformierter normalverteilter Zufallsvektoren zusammen. Wir verzichten dabei auf Beweise.

**Theorem 4.3** (Linear-affine Transformation eines normalverteilten Zufallsvektors).  $\xi \sim N(\mu, \Sigma)$  sei ein normalverteilter  $n$ -dimensionaler Zufallsvektor und es sei  $v := A\xi + b$  mit  $A \in \mathbb{R}^{m \times n}$  und  $b \in \mathbb{R}^m$ . Dann gilt

$$v \sim N(A\mu + b, A\Sigma A^T) \quad (4.11)$$

◦

Für einen Beweis verweisen wir auf Anderson (2003). Die linear-affine Transformation eines multivariat normalverteilten Zufallsvektors ergibt wieder einen multivariat normalverteilten Zufallsvektor. Die Parameter des resultierenden normalverteilten Zufallsvektors ergeben sich dabei aus den Parametern des ursprünglichen Zufallsvektors und den Transformationsparametern.

#### Beispiel

Als Beispiel betrachten wir die linear-affine Transformation eines bivariaten normalverteilten Zufallsvektors  $\xi$ . Es seien

$$\mu := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{und} \quad \Sigma := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix} \quad (4.12)$$

der Erwartungswert- und Kovarianzmatrixparameter von  $\xi$  und es seien

$$A := \begin{pmatrix} -2 & 1 \\ -1 & 2 \end{pmatrix} \quad \text{und} \quad b := \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (4.13)$$

die Transformationsparameter. Abbildung 4.5 A zeigt Isokonturen der WDF von  $\xi$  und Realisierungen  $x^{(i)} \in \mathbb{R}^2$  von  $\xi$  für  $i = 1, \dots, 50$ . Abbildung 4.5 B zeigt die transformierten Realisierungen  $y^{(i)} = Ax^{(i)} + b \in \mathbb{R}^2$  von  $v$  sowie die Isokonturen der WDF von  $v$  gemäß Theorem 4.4.

Als zweite lineare Transformation normalverteilter Zufallsvektoren wollen wir an dieser Stelle noch die invertierbare lineare Transformation betrachten. Diese ist Aussage des folgenden Theorems.

**Theorem 4.4** (Invertierbare lineare Transformation eines normalverteilten Zufallsvektors).  $\xi \sim N(\mu, \Sigma)$  sei ein normalverteilter  $n$ -dimensionaler Zufallsvektor und es sei  $v := A\xi$  mit einer invertierbaren Matrix  $A \in \mathbb{R}^{n \times n}$ . Dann gilt

$$v \sim N(A\mu, A\Sigma A^T) \quad (4.14)$$

◦

Offensichtlich handelt es sich hierbei um einen Spezialfall von Theorem 4.3 mit  $A \in \mathbb{R}^{n \times n}$  und  $b = 0_n$ . Die invertierbare lineare Transformation eines multivariat normalverteilten Zufallsvektors ergibt wieder einen multivariat normalverteilten Zufallsvektor. Die Parameter des resultierenden normalverteilten Zufallsvektors ergeben sich dabei aus den Parametern des ursprünglichen Zufallsvektors und der Transformationsmatrix.

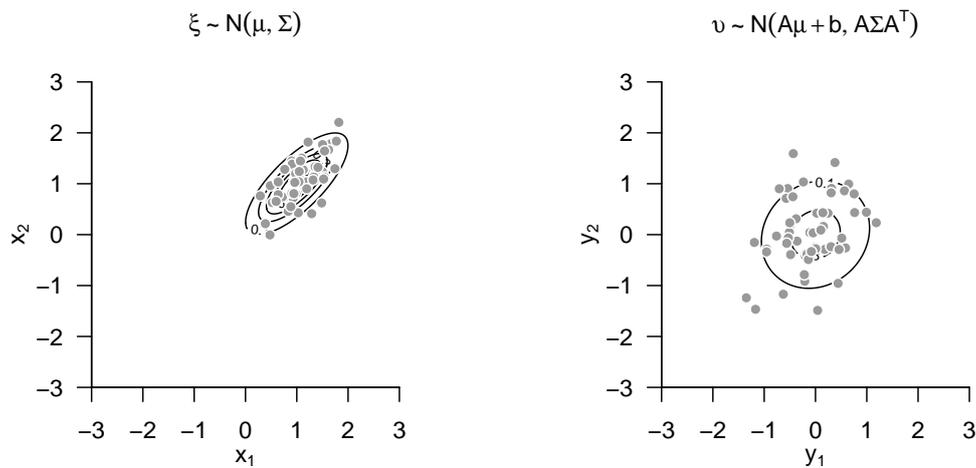


Abbildung 4.5 Linear-affine Transformation eines normalverteilten Zufallsvektors

## 4.4 Sphärische Verteilungen

Folgendes Theorem ist für die grundlegende Theorie des Allgemeinen Linearen Modells zentral.

**Theorem 4.5** (Sphärische multivariate Normalverteilung). Für  $i = 1, \dots, n$  seien  $N(x_i; \mu_i, \sigma^2)$  die WDFen von  $n$  unabhängigen univariaten normalverteilten Zufallsvariablen  $\xi_1, \dots, \xi_n$  mit  $\mu_1, \dots, \mu_n \in \mathbb{R}$  und  $\sigma^2 > 0$ . Weiterhin sei  $N(x; \mu, \sigma^2 I_n)$  die WDF eines  $n$ -dimensionalen Zufallsvektors  $\xi$  mit Erwartungswertparameter  $\mu := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$ . Dann gilt

$$p_\xi(x) = p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i) \quad (4.15)$$

und damit im vorliegenden Fall normalverteilter Zufallsvariablen

$$N(x; \mu, \sigma^2 I_n) = \prod_{i=1}^n N(x_i; \mu_i, \sigma^2). \quad (4.16)$$

◦

*Beweis.* Wir zeigen die Identität der multivariaten WDF  $N(x; \mu, \sigma^2 I_n)$  mit dem Produkt von  $n$  univaria-

ten WDFen  $N(x_i; \mu_i, \sigma^2)$ , wobei  $\mu_i$  der  $i$ te Eintrag von  $\mu \in \mathbb{R}^n$  ist. Es ergibt sich

$$\begin{aligned}
 N(x; \mu, \sigma^2 I_n) &= \frac{1}{\sqrt{(2\pi)^n |\sigma^2 I_n|}} \exp\left(-\frac{1}{2}(x - \mu)^\top (\sigma^2 I_n)^{-1} (x - \mu)\right) \\
 &= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}}\right) \frac{1}{\sqrt{(\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^\top (x - \mu)\right) \\
 &= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2\right) \\
 &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_i)^2\right) \\
 &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_i)^2\right) \\
 &= \prod_{i=1}^n N(x_i; \mu_i, \sigma^2).
 \end{aligned} \tag{4.17}$$

□

Einen Kovarianzmatrixparameter der Form  $\Sigma = \sigma^2 I_n$  nennt man auch *sphärisch*, da die Isokonturen der WDF eines normalverteilten Zufallsvektors mit einem solchen Kovarianzmatrixparameter *Sphären* bilden (zum Beispiel Kreise bei  $n = 2$  und Kugeln bei  $n = 3$ ). Eine multivariate Normalverteilung mit sphärische Kovarianzmatrixparameter nennt man entsprechend eine *sphärische Normalverteilung*. Theorem 4.5 besagt, dass die WDF eines  $n$ -dimensionalen normalverteilten Zufallsvektors mit sphärischem Kovarianzparameter der gemeinsamen WDF von  $n$  unabhängigen univariat normalverteilten Zufallsvariablen entspricht und umgekehrt. Eine Realisierung eines  $n$ -dimensionalen normalverteilten Zufallsvektors entspricht also den Realisierungen von  $n$  unabhängigen univariat normalverteilten Zufallsvariablen und umgekehrt. Man beachte, dass die Identität der Verteilungen der  $\xi_i, i = 1, \dots, n$  hier nicht vorausgesetzt ist, insbesondere können sich ihre Erwartungswertparameter  $\mu_i, i = 1, \dots, n$  explizit unterscheiden.

## 4.5 Marginale und bedingte Verteilungen

Multivariate Normalverteilungen haben die Eigenschaft, dass auch alle anderen assoziierten Verteilungen Normalverteilungen sind und deren Erwartungswert- und Kovarianzmatrixparameter aus den Parametern der jeweils komplementären Verteilung errechnet werden können. Insbesondere gilt zum einen, dass die marginalen und bedingten Verteilungen multivariater Normalverteilungen wiederum Normalverteilungen sind. Zum anderen lassen sich multivariate Normalverteilungen wie alle multivariaten Verteilungen multiplikativ in eine marginale und eine bedingte Verteilung zerlegen. Insbesondere sind nun aber bei multivariaten Normalverteilungen diese Verteilungen wiederum (multivariate) Normalverteilungen, deren Parameter aus den Parametern der gemeinsame Verteilung errechnet werden können und umgekehrt. Wir fassen die obigen Erkenntnisse formal in den folgenden drei Theoremen zusammen.

**Theorem 4.6** (Marginale Normalverteilungen). *Es sei  $m := k + l$  und  $\xi = (\xi_1, \dots, \xi_m)^\top$  sei ein  $m$ -dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter*

$$\mu = \begin{pmatrix} \mu_\nu \\ \mu_\zeta \end{pmatrix} \in \mathbb{R}^m, \tag{4.18}$$

mit  $\mu_v \in \mathbb{R}^k$  and  $\mu_\zeta \in \mathbb{R}^l$  und Kovarianzmatrixparameter

$$\Sigma = \begin{pmatrix} \Sigma_{vv} & \Sigma_{v\zeta} \\ \Sigma_{\zeta v} & \Sigma_{\zeta\zeta} \end{pmatrix} \in \mathbb{R}^{m \times m}, \tag{4.19}$$

mit  $\Sigma_{vv} \in \mathbb{R}^{k \times k}$ ,  $\Sigma_{v\zeta} \in \mathbb{R}^{k \times l}$ ,  $\Sigma_{\zeta v} \in \mathbb{R}^{l \times k}$ , und  $\Sigma_{\zeta\zeta} \in \mathbb{R}^{l \times l}$ . Dann sind  $v := (\xi_1, \dots, \xi_k)^T$  und  $\zeta := (\xi_{k+1}, \dots, \xi_m)^T$   $k$ - bzw.  $l$ -dimensionale normalverteilte Zufallsvektoren und es gilt:

$$v \sim N(\mu_v, \Sigma_{vv}) \text{ and } \zeta \sim N(\mu_\zeta, \Sigma_{\zeta\zeta}). \tag{4.20}$$

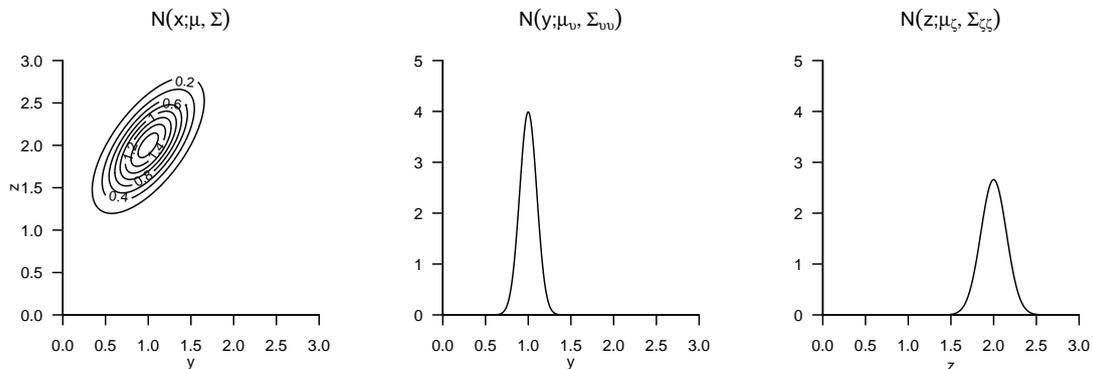
◦

Die Marginalverteilungen einer multivariaten Normalverteilung sind also auch Normalverteilungen und die Parameter der Marginalverteilungen ergeben sich aus den Parametern der gemeinsamen Verteilung. Für Beweise dieses Theorems verweisen wir auf Mardia, Kent, und Bibby (1979) und Anderson (2003).

Abbildung 4.6 visualisiert Theorem 4.6 für den Fall  $n := 2$ ,  $k := 1$ ,  $l := 1$ ,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2 \text{ und } \Sigma := \begin{pmatrix} 0.10 & 0.08 \\ 0.08 & 0.15 \end{pmatrix} \in \mathbb{R}^{2 \times 2}. \tag{4.21}$$

Abbildung 4.6 A zeigt dabei die WDF des bivariaten Zufallsvektors  $\xi$  und Abbildung 4.6 B und C zeigen die WDFen der entsprechenden marginalen Zufallsvariablen  $v$  und  $\zeta$ .



**Abbildung 4.6** Marginale Verteilungen eines bivariaten normalverteilten Zufallsvektor.

Die Definition einer multivariaten Normalverteilung erlaubt es weiterhin, die bedingten Verteilungen aller Komponenten des entsprechenden Zufallsvektors direkt mithilfe der Parameter der multivariaten Normalverteilung zu bestimmen. Dies ist die zentrale Aussage folgenden Theorems.

**Theorem 4.7** (Bedingte Normalverteilungen).  $\begin{pmatrix} \xi \\ v \end{pmatrix}$  sei ein  $(m+n)$ -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_{\xi,v} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p_{\xi,v} \left( \begin{pmatrix} x \\ y \end{pmatrix} \right) := N \left( \begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi,v}, \Sigma_{\xi,v} \right) \quad (4.22)$$

mit

$$\mu_{\xi,v} = \begin{pmatrix} \mu_{\xi} \\ \mu_v \end{pmatrix} \text{ und } \Sigma_{\xi,v} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi v} \\ \Sigma_{v\xi} & \Sigma_{vv} \end{pmatrix}, \quad (4.23)$$

wobei  $x, \mu_{\xi} \in \mathbb{R}^m, y, \mu_v \in \mathbb{R}^n$  und  $\Sigma_{\xi\xi} \in \mathbb{R}^{m \times m}, \Sigma_{\xi v} \in \mathbb{R}^{m \times n}, \Sigma_{vv} \in \mathbb{R}^{n \times n}$ . Dann ist die bedingte Verteilung von  $\xi$  gegeben  $v$  eine  $m$ -dimensionale Normalverteilung mit bedingter WDF

$$p_{\xi|v}(\cdot|y) : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_{\xi|v}(x|y) := N(x; \mu_{\xi|v}, \Sigma_{\xi|v}) \quad (4.24)$$

mit

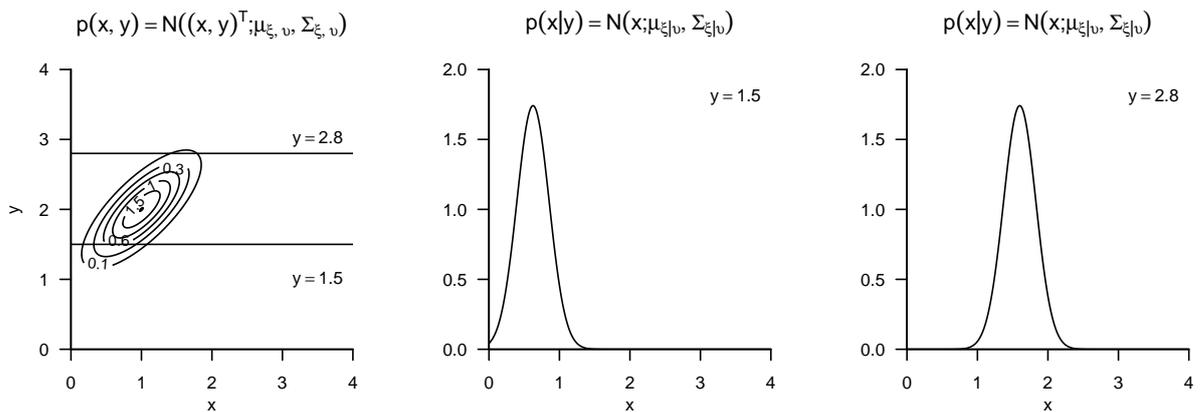
$$\mu_{\xi|v} = \mu_{\xi} + \Sigma_{\xi v} \Sigma_{vv}^{-1} (y - \mu_v) \in \mathbb{R}^m \quad (4.25)$$

und

$$\Sigma_{\xi|v} = \Sigma_{\xi\xi} - \Sigma_{\xi v} \Sigma_{vv}^{-1} \Sigma_{v\xi} \in \mathbb{R}^{m \times m}. \quad (4.26)$$

◦

Die bedingten Verteilungen einer multivariaten Normalverteilung sind also auch Normalverteilungen und die Parameter der bedingten Verteilungen ergeben sich aus den Parametern der gemeinsamen Verteilung.



**Abbildung 4.7** Bedingte Verteilungen eines bivariaten normalverteilten Zufallsvektor.

Abbildung 4.7 visualisiert Theorem 4.7 für den Fall  $m := 2, n := 1$ ,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 0.12 & 0.09 \\ 0.09 & 0.12 \end{pmatrix}. \quad (4.27)$$

Abbildung 4.7 A zeigt dabei die WDF des bivariaten Zufallsvektors  $(\xi, v)^T$  und Abbildung 4.7 B und C zeigen die WDF der bedingten Verteilung der Zufallsvariable  $\xi$  gegeben  $v = 1.5$  bzw.  $v = 2.8$ .

Mithilfe einer marginalen und einer bedingten multivariaten Normalverteilung lässt sich eine gemeinsame multivariate Normalverteilung konstruieren, deren Parameter sich aus den Parametern der marginalen und bedingten Verteilung ergeben. Dies ist die zentrale Aussage folgenden Theorems.

**Theorem 4.8** (Gemeinsame Normalverteilungen).  $\xi$  sei ein  $m$ -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_\xi : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_\xi(x) := N(x; \mu_\xi, \Sigma_{\xi\xi}) \text{ mit } \mu_\xi \in \mathbb{R}^m, \Sigma_{\xi\xi} \in \mathbb{R}^{m \times m}, \quad (4.28)$$

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

$$p_{v|\xi}(\cdot|x) : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, y \mapsto p_{v|\xi}(y|x) := N(y; A\xi + b, \Sigma_{vv}) \text{ mit } \Sigma_{vv} \in \mathbb{R}^{n \times n}. \quad (4.29)$$

Dann ist der  $(m+n)$ -dimensionale Zufallsvektor  $\begin{pmatrix} \xi \\ v \end{pmatrix}$  normalverteilt mit (gemeinsamer) WDF

$$p_{\xi,v} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p_{\xi,v} \left( \begin{pmatrix} x \\ y \end{pmatrix} \right) = N \left( \begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi,v}, \Sigma_{\xi,v} \right), \quad (4.30)$$

mit  $\mu_{\xi,v} \in \mathbb{R}^{m+n}$  and  $\Sigma_{\xi,v} \in \mathbb{R}^{(m+n) \times (m+n)}$  gegeben durch

$$\mu_{\xi,v} = \begin{pmatrix} \mu_\xi \\ A\mu_\xi + b \end{pmatrix} \text{ und } \Sigma_{\xi,v} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\xi}A^T \\ A\Sigma_{\xi\xi} & \Sigma_{vv} + A\Sigma_{\xi\xi}A^T \end{pmatrix}. \quad (4.31)$$

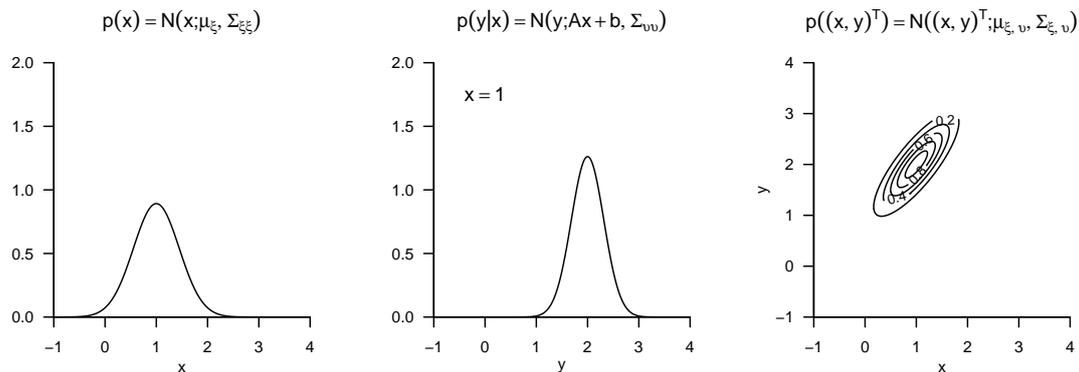
◦

Insbesondere ergeben sich die Parameter der gemeinsamen Verteilung also als linear-affine Transformation der Parameter der induzierenden Verteilungen.

Abbildung 4.8 visualisiert Theorem 4.8 für den Fall  $m := 1$ ,  $n := 1$ ,  $\mu_\xi := 1$ ,  $\Sigma_{\xi\xi} := 0.2$ ,  $A := 1$ ,  $b := 1$  und  $\Sigma_{vv} := 0.1$ . Abbildung 4.8 A zeigt dabei die WDF der Zufallsvariable  $\xi$ , Abbildung 4.8 B zeigt die WDF der bedingten Verteilung der Zufallsvariable  $v$  gegeben  $\xi$  und Abbildung 4.6 C schließlich zeigt die WDF des induzierten bivariaten Zufallsvektors  $(\xi, v)$ .

## 4.6 Literaturhinweise

Die Entwicklung der bivariaten Normalverteilung hat ihre Ursprünge in der statistischen Literatur zur Mitte des 19. Jahrhunderts, insbesondere in den Arbeiten von Francis Galton (1822-1911). Die mathematische Formalisierung der bivariaten Normalverteilung geht dabei wohl auf Pearson (1896) zurück (siehe Seal (1967)). Die ursprüngliche Formulierung der multivariaten Normalverteilung wird bei Edgeworth (1892) verortet. Tong (1990) gibt eine umfassenden Überblick zur Theorie und Anwendung der multivariaten Normalverteilung.



**Abbildung 4.8** Gemeinsame Verteilungen einer marginalen und einer auf dieser bedingten normalverteilten Zufallsvariable.

## 4.7 Selbstkontrollfragen

1. Geben Sie die Definitionen des Erwartungswerts und der Kovarianzmatrix eines Zufallsvektors wieder.
2. Was repräsentieren die Diagonalelemente der Kovarianzmatrix eines Zufallsvektors?
3. Was repräsentieren die Nichtdiagonalelemente der Kovarianzmatrix eines Zufallsvektors?
4. Geben Sie die Definition eines multivariat normalverteilten Zufallsvektors wieder.
5. Erläutern Sie die Definition eines multivariat normalverteilten Zufallsvektors.
6. Welche Werte haben der Erwartungswert und die Kovarianzmatrix eines normalverteilten Zufallsvektors?
7. Visualisieren Sie die WDF eines 2-dimensionalen normalverteilten Zufallsvektors mit den Parameterwerten

$$\mu := \begin{pmatrix} 10 \\ 15 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}. \quad (4.32)$$

8. Generieren Sie 100 Realisierungen aus dieser Verteilung und visualisieren Sie diese.
9. Geben Sie das Theorem zur linear-affinen Transformation multivariater Normalverteilungen wieder.
10. Geben Sie das Theorem zur invertierbaren linearen Transformation multivariater Normalverteilungen wieder.
11. Geben Sie das Theorem zu sphärischen Normalverteilungen wieder.
12. Erläutern Sie den Begriff des sphärischen Kovarianzmatrixparameters.
13. Skizzieren Sie den Beweis des Theorems zu sphärischen Normalverteilungen.

# 5 Modellformulierung

## 5.1 Allgemeine Theorie

Wir definieren das Allgemeine Lineare Modell (ALM) wie folgt.

**Definition 5.1** (Allgemeines Lineares Modell).

Es sei

$$y = X\beta + \varepsilon \quad (5.1)$$

wobei

- $y$  ein  $n$ -dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird,
- $X \in \mathbb{R}^{n \times p}$  für  $n > p$  eine vorgegebene Matrix ist, die *Designmatrix* genannt wird,
- $\beta \in \mathbb{R}^p$  ein unbekannter Parametervektor ist, der *Betaparametervektor* genannt wird,
- $\varepsilon$  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

$$\varepsilon \sim N(0_n, \sigma^2 I_n) \quad (5.2)$$

Dann heißt Gleichung 5.1 *Allgemeines Lineares Modell (ALM)*.

•

Wir nehmen durchgängig an, dass  $X \in \mathbb{R}^{n \times p}$  vollen Spaltenrang hat, dass also  $\text{rg}(X) = p$  ist. In Gleichung 5.1 bezeichnen wir  $X\beta \in \mathbb{R}^n$  als den *deterministischen Aspekt* des ALMs und  $\varepsilon$  als den *probabilistischen Aspekt* des ALMs. Das ALM postuliert also, dass Daten aus der Addition eines deterministischen Aspektes  $X\beta$  unter der Addition eines multivariat normalverteilten probabilistischen Aspektes  $\varepsilon$  zustande kommen. Man beachte, dass  $X\beta$  ein  $n$ -dimensionaler Vektor und  $\varepsilon$  ein  $n$ -dimensionaler Zufallsvektor ist. Der resultierende Vektor  $y$  ist ein Zufallsvektor, weil er aus der Addition des Zufallsvektors  $\varepsilon$  zu dem konstanten Vektor  $X\beta$  resultiert. Das ALM ist also ein probabilistisches Modell, bei dem durch  $y \in \mathbb{R}^n$  vorliegende Datensätze modelliert werden. Generativ betrachtet, entsteht im ALM ein Datensatz  $y$  dann durch Addition des deterministischen Modellaspekts und einer nicht direkt beobachtbaren Realisierung  $\varepsilon$ ,

$$y = X\beta + \varepsilon \quad (5.3)$$

mit  $y \in \mathbb{R}^n$ ,  $X\beta \in \mathbb{R}^n$  und  $\varepsilon \in \mathbb{R}^n$ .

Die Gesamtzahl an Parametern des ALMs beträgt  $p + 1$ , bestehend aus  $p$  skalaren Betaparametern und einem Varianzparameter  $\sigma^2$ . Der Betaparametervektor  $\beta \in \mathbb{R}^p$  wird dabei auch Gewichtsvektor oder Effektvektor genannt. Seine Einträge wichten die Einträge der Spalten der Designmatrix  $X \in \mathbb{R}^{n \times p}$  in der Erzeugung des deterministischen

Modellaspekts  $X\beta \in \mathbb{R}^n$ . Die Spalten der Designmatrix werden in unterschiedlichen Kontexten unterschiedlich bezeichnet. Gebräuchliche Bezeichnungen sind zum Beispiel *Prädiktoren*, *Regressoren* oder *Kovariaten*. Allgemein betrachtet, modellieren die Spalten der Designmatrix unabhängige Variablen und der Datenvektor abhängige Variablen.

Man beachte, dass der Kovarianzmatrixparameter von  $\varepsilon$  als sphärisch angenommen wird. Damit folgt direkt, dass die  $\varepsilon_1, \dots, \varepsilon_n$  unabhängige normalverteilte Zufallsvariablen mit identischem Varianzparameter sind. Weil für  $\varepsilon$  zusätzlich der Erwartungswertparameter als  $0_n \in \mathbb{R}^n$  angenommen wird, sind die  $\varepsilon_1, \dots, \varepsilon_n$  auch identisch normalverteilte Zufallsvariablen.

Wenn  $x_{ij} \in \mathbb{R}$  das  $ij$ te Element der Designmatrix  $X \in \mathbb{R}^{n \times p}$  bezeichnet, dann gilt damit für jede Komponente  $y_i, i = 1, \dots, n$  von  $y$  nach Gleichung 5.1, dass

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i \text{ mit } \varepsilon_1, \dots, \varepsilon_n \sim N(0, \sigma^2). \quad (5.4)$$

Die in Gleichung 5.1 implizite Verteilung des Datenvektors  $y$  halten wir in folgendem Theorem fest.

**Theorem 5.1** (Datenverteilung des Allgemeinen Linearen Modells).

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (5.5)$$

das ALM. Dann gilt

$$y \sim N(\mu, \sigma^2 I_n) \text{ mit } \mu := X\beta \in \mathbb{R}^n. \quad (5.6)$$

◦

*Beweis.* Das Theorem zur linear-affinen Transformation multivariater Normalverteilungen besagt:

$$\xi \in \mathbb{R}^n, \xi \sim N(\mu, \Sigma), A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \Rightarrow v := A\xi + b \sim N(A\mu + b, A\Sigma A^T). \quad (5.7)$$

Damit gilt für  $\varepsilon \sim N(0_n, \sigma^2 I_n)$  und  $y := I_n \varepsilon + X\beta$ , dass

$$y \sim N(I_n 0_n + X\beta, I_n(\sigma^2 I_n)I_n^T) = N(X\beta, \sigma^2 I_n) = N(\mu, \sigma^2 I_n) \text{ mit } \mu := X\beta \in \mathbb{R}^n. \quad (5.8)$$

□

Im ALM sind die Daten  $y$  also ein  $n$ -dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter  $\mu = X\beta \in \mathbb{R}^n$  und Kovarianzmatrixparameter  $\sigma^2 I_n \in \mathbb{R}^{n \times n}$ . Das ALM ist also eine multivariate Normalverteilung, deren Erwartungswertparameter mithilfe einer Designmatrix und eines Betaparametervektors parameterisiert ist. Weiterhin sind die Komponenten  $y_1, \dots, y_n$  von  $y$ , also die Zufallsvariablen, die skalare Datenpunkte modellieren, damit unabhängige normalverteilte Zufallsvariablen der Form

$$y_i \sim N((X\beta)_i, \sigma^2) \text{ für } i = 1, \dots, n. \quad (5.9)$$

Da im Allgemeinen aber  $(X\beta)_i \neq (X\beta)_j$  für  $i \neq j$  gilt, sind die  $y_i, i = 1, \dots, n$  in den meisten Fällen nicht identisch verteilt. Das Szenario unabhängig und identisch normalverteilter Zufallsvariablen kann natürlich dennoch als Spezialfall des ALMs formuliert werden.

### Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

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

$$y_1, \dots, y_n \sim N(\mu, \sigma^2). \quad (5.10)$$

Dann gilt, dass Gleichung 5.10 äquivalent ist zu

$$y_i = \mu + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \text{ mit unabhängigen } \varepsilon_i. \quad (5.11)$$

In Matrixschreibweise ist dies wiederum äquivalent zu dem ALM-Spezialfall

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}^1, \sigma^2 > 0. \quad (5.12)$$

In **R** können Realisierungen des ALMs leicht mithilfe eines Zufallszahlengenerators der multivariaten Normalverteilung durch Spezifikation der entsprechenden Erwartungswert- und Kovarianzmatrixparameter gewonnen werden. Folgender **R**-Code zeigt, wie  $n$  unabhängig und identisch normalverteilte skalare Datenpunkte im Sinne des ALMs realisiert werden können. Man beachte, dass  $n$  skalare Datenpunkte dabei einer Realisierung des ALMs entsprechen.

```
# Modellformulierung
library(MASS) # multivariate Normalverteilung
set.seed(0) # Zufallszahlengenerator initialisieren
n = 12 # Anzahl von Datenpunkten
p = 1 # Anzahl von Betaparametern
X = matrix(rep(1,n), nrow = n) # n x p 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 des n-dimensionalen ZVs y
```

Realisierungen: 1.2 2.76 4.4 1.99 1.71 1.07 0.46 2.41 3.27 3.33 1.67 3.26

### Beispiel (2) Einfache lineare Regression

Wir betrachten das generative Modell der einfachen linearen Regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n. \quad (5.13)$$

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

$$y_i \sim N(\mu_i, \sigma^2) \text{ mit } \mu_i := \beta_0 + \beta_1 x_i \text{ für } i = 1, \dots, n. \quad (5.14)$$

In Matrixschreibweise ist dies wiederum äquivalent zu dem ALM-Spezialfall

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^2, \sigma^2 > 0. \quad (5.15)$$

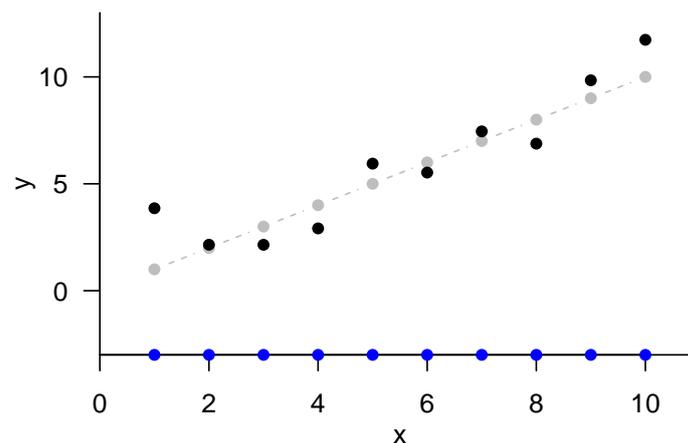
**R**-Code zur Simulation von Realisierungen einer einfachen linearen Regression hat dementsprechend eine sehr ähnliche Struktur wie obiger **R**-Code zur Simulation von Realisierungen von  $n$  unabhängig und identisch normalverteilten Zufallsvariablen.

```
# Modellformulierung
library(MASS) # multivariate Normalverteilung
set.seed(0) # Zufallzahlengenerator initialisieren
n = 10 # Anzahl von Datenpunkten
p = 2 # Anzahl von Betaparametern
x = 1:n # Prädiktorwerte
X = matrix(c(rep(1,n),x), nrow = n) # n x p 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 des n-dimensionalen ZVs y
```

Realisierungen: 3.4 1.99 2.71 3.07 3.46 6.41 8.27 9.33 8.67 11.26

Wir visualisieren diese Realisierung in Abbildung 5.1.



**Abbildung 5.1** Realisierung des Modells der einfachen linearen Regression.

## 5.2 Identifizierbarkeit und Schätzbarkeit

Wir haben oben gesehen, dass die Datenverteilung des ALMs durch

$$y \sim N(\mu, \sigma^2 I_n) \text{ mit } \mu := X\beta \in \mathbb{R}^n \text{ für } X \in \mathbb{R}^{n \times p}, \beta \in \mathbb{R}^p \quad (5.16)$$

gegeben ist. Das ALM ist also eine multivariate Normalverteilung mit einer speziellen Erwartungswertparameterisierung. Um die Begriffe der Identifizierbarkeit und Schätzbarkeit im Kontext von ALMs einzuführen, ist es hilfreich, den Begriff der *Parametrisierung einer multivariaten Normalverteilung* zunächst etwas zu verallgemeinern.

**Definition 5.2** (Betaparametrisierung).  $N(\mu, \sigma^2 I_n)$  sei eine multivariate Normalverteilung mit sphärischem Kovarianzmatrixparameter. Dann bezeichnen wir eine multivariate, vektorwertige Funktion der Form

$$f : \mathbb{R}^p \rightarrow \mathbb{R}^n, \beta \mapsto f(\beta) := \mu \quad (5.17)$$

als eine Betaparametrisierung von  $\mu$ .

•

Das ALM beruht offenbar auf der Designmatrix-abhängigen Betaparametrisierung

$$f : \mathbb{R}^p \rightarrow \mathbb{R}^n, \beta \mapsto f(\beta) := X\beta. \quad (5.18)$$

Mit Hilfe des Begriffs der Betaparametrisierung können wir nun den Begriff der *Betaparameteridentifizierbarkeit* formulieren:

**Definition 5.3** (Betaparameteridentifizierbarkeit).  $N(\mu, \sigma^2 I_n)$  sei eine multivariate Normalverteilung mit sphärischem Kovarianzmatrixparameter und  $f$  sei eine Betaparametrisierung von  $\mu$ .  $\beta$  heißt dann und nur dann identifizierbar, wenn für beliebige  $\beta_1, \beta_2 \in \mathbb{R}^p$  gilt, dass aus  $f(\beta_1) = f(\beta_2)$  folgt, dass  $\beta_1 = \beta_2$  gilt.

•

Die Betaparameter allgemeiner linearer Modelle, deren Designmatrix vollen Rang hat, sind identifizierbar. Dies ist die Aussage folgenden Theorems:

**Theorem 5.2** (Betaparameteridentifizierbarkeit bei vollem Spaltenrang der Designmatrix).  $y \sim N(X\beta, \sigma^2 I_n)$  sei die Datenverteilung eines ALMs mit  $\text{rg}(X) = p$ . Dann ist  $\beta$  identifizierbar.

◦

*Beweis.* Für  $X \in \mathbb{R}^{n \times p}$  impliziert  $\text{rg}(X) = p$ , dass  $(X^T X) \in \mathbb{R}^{p \times p}$  eine invertierbare Matrix ist. Dann aber gilt für beliebige  $\beta_1, \beta_2 \in \mathbb{R}^p$

$$\begin{aligned} f(\beta_1) &= f(\beta_2) \\ \Leftrightarrow X\beta_1 &= X\beta_2 \\ \Leftrightarrow X^T X\beta_1 &= X^T X\beta_2 \\ \Leftrightarrow (X^T X)^{-1} X^T X\beta_1 &= (X^T X)^{-1} X^T X\beta_2 \\ \Leftrightarrow \beta_1 &= \beta_2. \end{aligned} \quad (5.19)$$

□

Im Rahmen der Analyse schätzbarer Funktionen benötigen wir weiterhin den Begriff der Identifizierbarkeit von vektorwertigen Funktionen der Betaparameter. Wir definieren:

**Definition 5.4** (Identifizierbarkeit von Funktionen der Betaparameter).  $N(\mu, \sigma^2 I_n)$  sei eine multivariate Normalverteilung mit sphärischem Kovarianzmatrixparameter und  $f$  sei eine Betaparametrisierung von  $\mu$ . Weiterhin sei

$$g : \mathbb{R}^p \rightarrow \mathbb{R}^k, \beta \mapsto g(\beta) \quad (5.20)$$

eine Funktion des Betaparametervektors.  $g$  heißt dann und nur dann identifizierbar, wenn für beliebige  $f(\beta_1), f(\beta_2) \in \mathbb{R}^n$  gilt, dass aus  $f(\beta_1) = f(\beta_2)$  folgt, dass  $g(\beta_1) = g(\beta_2)$ .

•

*Schätzbare Funktionen* sind lineare Funktionen von  $\beta$ , die identifizierbar sind. Allgemein gilt folgendes Theorem:

**Theorem 5.3** (Identifizierbare Betaparameterfunktionen).  $N(\mu, \sigma^2 I_n)$  sei eine multivariate Normalverteilung mit sphärischem Kovarianzmatrixparameter und  $f$  sei eine Betaparametrisierung von  $\mu$ . Weiterhin sei

$$g : \mathbb{R}^p \rightarrow \mathbb{R}^k, \beta \mapsto g(\beta) \quad (5.21)$$

eine Funktion des Betaparametervektors. Die Funktion  $g$  ist dann und nur dann identifizierbar, wenn  $g$  eine Funktion von  $f$  ist, wenn also eine Funktion  $\phi$  existiert, so dass

$$g = \phi \circ f \quad (5.22)$$

◦

*Beweis.* Wir zeigen die Aussage lediglich in eine Richtung. Wir nehmen also an, es existiert eine Funktion  $\phi$ , so dass

$$g = \phi \circ f. \quad (5.23)$$

Dann impliziert die Tatsache, dass eine Funktion einem Argument genau einen Funktionswert zuordnet, dass gilt

$$f(\beta_1) = f(\beta_2) \Leftrightarrow \phi(f(\beta_1)) = \phi(f(\beta_2)) \Leftrightarrow g(\beta_1) = g(\beta_2) \quad (5.24)$$

Also ist  $g$  identifizierbar, denn aus  $f(\beta_1) = f(\beta_2)$  folgt, dass  $g(\beta_1) = g(\beta_2)$ . □

Wie oben bereits erwähnt, sind schätzbare Funktionen lineare Funktionen von  $\beta$ , die identifizierbar sind. Die klassische Definition einer schätzbaren Funktion ist folgende.

**Definition 5.5** (Schätzbare Funktion).  $N(X\beta, \sigma^2 I_n)$  sei ein ALM. Dann heißt eine lineare Funktion

$$g : \mathbb{R}^p \rightarrow \mathbb{R}^k, \beta \mapsto g(\beta) := C^T \beta \quad (5.25)$$

mit  $C \in \mathbb{R}^{p \times k}$  schätzbare, wenn eine Matrix  $P \in \mathbb{R}^{n \times k}$  existiert, sodass

$$C^T \beta = P^T X \beta. \quad (5.26)$$

•

Diese Definition erschließt sich wie folgt: Nach dem Theorem zu identifizierbaren Funktionen muss eine identifizierbare lineare Funktion von  $\beta$  eine Funktion der Form

$$g(\beta) = (\phi \circ f)(\beta) \quad (5.27)$$

sein. Da weiterhin gilt, dass für ein ALM  $f(\beta) = X\beta$  und dass  $g$  eine lineare Funktion ist, also mithilfe einer Matrix  $C^T$  geschrieben werden kann, muss auch  $\phi$  eine lineare Funktion sein. Damit kann aber auch  $\phi$  geschrieben werden als

$$\phi(f(\beta)) = \phi(X\beta) = P^T X\beta \quad (5.28)$$

mit einer geeigneten Matrix  $P \in \mathbb{R}^{n \times k}$ .

### 5.3 Designspektrum

Die Wahl von Designmatrix und Betaparameter eröffnet eine große Freiheit zur Implementation verschiedenster Erwartungswertparameterszenarien der ALM-Datenverteilung. Prinzipiell liegen dabei alle speziellen Designs in einem Kontinuum zwischen folgenden beiden Extrema:

- (1) Die Erwartungswerte aller Datenvariablen sind identisch, d.h.

$$y_i \sim N(\mu, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n, \quad (5.29)$$

also

$$y = X\beta + \varepsilon \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}, \varepsilon \sim N(0_n, \sigma^2 I_n). \quad (5.30)$$

- (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden, d.h.

$$y_i \sim N(\mu_i, \sigma^2) \text{ u.v. für } i = 1, \dots, n, \quad (5.31)$$

also

$$y = X\beta + \varepsilon \text{ mit } X := I_n \in \mathbb{R}^{n \times n}, \beta := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n, \varepsilon \sim N(0_n, \sigma^2 I_n). \quad (5.32)$$

In Szenario (1) wird jegliche Datenvariabilität dem Zufallsfehlerterm zugeschrieben, in Szenario (2) wird dagegen jegliche Datenvariabilität dem Erwartungswertparameter zugeschrieben. Beide Extremszenarien sind wissenschaftlich nicht ergiebig, da sie keine theoriegeleitete systematische Abhängigkeit zwischen unabhängigen und abhängigen Variablen anzunehmen im Stande sind. Alle im weiteren Verlauf betrachteten ALM-Designs liegen damit zwischen den beiden Extremszenarien und repräsentieren verschiedene Formen der systematischen Abhängigkeit zwischen unabhängigen und abhängigen Variablen. Insbesondere unterscheidet man:

- (1) *Faktorielle Designs*, bei denen die Designmatrix im Wesentlichen nur 1en und 0en (und manchmal  $-1$ en) enthält. In diesem Fall repräsentieren die Betaparameter Gruppenerwartungswerte und wir werden sehen, dass die Betaparameterschätzer aus (aus der Kombination von) Gruppenstichprobenmitteln bestehen. Man sagt manchmal, dass faktorielle Designs der Untersuchung von *Unterschiedshypothesen* dienen. Beispiele für faktorielle Designs sind verschiedene Designs zur Implementation von *T-Tests* und *Varianzanalysen*.

- (2) *Parameterische Designs*, bei denen die Designmatrix aus Spalten mit kontinuierlichen reellen Werten besteht. Vor allem in diesem Kontext werden die Spalten der Designmatrix oft als *Regressoren* oder *Prädiktoren* bezeichnet. In diesem Fall repräsentieren die Betaparameter partielle Steigungsparameter und wir werden sehen, dass sich die entsprechenden Betaparameterschätzer als normalisierte Regressor-Daten-Kovarianzen ergeben. Man sagt manchmal, dass parametrische Designs zur Untersuchung von *Zusammenhangshypothesen* dienen. Beispiele für parametrische Designs sind verschiedene Designs zur Implementation von *einfacher linearer Regression* und insbesondere *multipler linearer Regression*.
- (3) *Faktoriell-parametrische Designs*, bei denen die Spalten der Designmatrix sowohl faktorielle als auch parametrische Prädiktoren repräsentieren. In diesem Kontext werden die parametrischen Regressoren oft als *Kovariaten* betrachtet. Gemischt faktoriell-parametrische Designs sind das zentrale Charakteristikum der *Kovarianzanalyse* die auf eine kontrollierte Untersuchung von Unterschiedshypothesen bei Vorliegen weiterer möglicher Abhängigkeiten zwischen unabhängigen und abhängigen Variablen bzw. auf die kontrollierte Untersuchung von Zusammenhangshypothesen bei Vorliegen weiterer möglicher Gruppenunterschiede abzielt.

## 5.4 Literaturhinweise

Das Allgemeine Lineare Modell hat eine lange Geschichte, deren moderne Inkarnation üblicherweise auf die Arbeiten von Legendre (1805) und Gauss (1809) zurückgeführt wird. Matrixbasierte Formulierungen der multiplen Regression finden sich spätestens bei Aitken (1936) und Scheffé (1959). Eingang in den psychologischen Methodenkanon findet das Allgemeine Lineare Modell spätestens mit Cohen (1968). Seal (1967) gibt einen ausführlichen Überblick zur Geschichte des Allgemeinen Linearen Modells im 19. und in der ersten Hälfte des 20. Jahrhunderts. Die in diesem Abschnitt gegebene Diskussion von Identifizierbarkeit und Schätzbarkeit beruht auf der Darstellung in Christensen (2011).

## 5.5 Selbstkontrollfragen

1. Erläutern Sie das naturwissenschaftliche Paradigma.
2. Erläutern Sie die Standardprobleme der Frequentistischen Inferenz.
3. Geben Sie die Definition des Allgemeinen Linearen Modells wieder.
4. Erläutern Sie die deterministischen und probabilistischen Aspekte des ALMs.
5. Wieviele skalare Parameter hat das ALM mit sphärischer Kovarianzmatrix?
6. Warum sind die Komponenten des ALM Zufallsfehlers unabhängig und identisch verteilt?
7. Geben Sie das Theorem zur Datenverteilung des Allgemeinen Linearen Modells wieder.
8. Sind die Komponenten des ALM-Datenvektors immer unabhängig und identisch verteilt?
9. Schreiben Sie das Szenario von  $n$  unabhängig und identisch normalverteilten Zufallsvariablen in ALM-Form.
10. Schreiben Sie das Szenario der einfachen linearen Regression in ALM-Form.

# 6 Parameterschätzung

In diesem Abschnitt betrachten wir die Frequentistische Punktschätzung von Betaparametervektor und Varianzparameter im ALM. Als Beispielanwendungen betrachten wir das Szenario  $n$  unabhängig und identisch normalverteilter Zufallsvariablen und das Szenario der einfachen linearen Regression. Wir schließen mit der Dokumentation der Frequentistischen Verteilungen der Parameterschätzer im Rahmen des ALMs.

## 6.1 Betaparameterschätzung

Wir fassen die Frequentistische Punktschätzung des Betaparametervektors in folgendem Theorem zusammen.

**Theorem 6.1** (Betaparameterschätzer). *Gegeben seien das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.1)$$

und der Betaparameterschätzer

$$\hat{\beta} := (X^T X)^{-1} X^T y. \quad (6.2)$$

Dann gilt, dass  $\hat{\beta}$  die Summe der Abweichungsquadrate minimiert,

$$\hat{\beta} = \arg \min_{\tilde{\beta}} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (6.3)$$

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

◦

*Beweis.*

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

$$(y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (6.4)$$

minimiert. Dazu halten wir zunächst fest, dass

$$\hat{\beta} = (X^T X)^{-1} X^T y \Leftrightarrow X^T X \hat{\beta} = X^T y \Leftrightarrow X^T y - X^T X \hat{\beta} = 0_p \Leftrightarrow X^T (y - X \hat{\beta}) = 0_p. \quad (6.5)$$

Weiterhin gilt dann auch, dass

$$X^T (y - X \hat{\beta}) = 0_p \Leftrightarrow (X^T (y - X \hat{\beta}))^T = 0_p^T \Leftrightarrow (y - X \hat{\beta})^T X = 0_p^T. \quad (6.6)$$

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

$$z^T X^T X z \geq 0 \quad \text{für alle } z \in \mathbb{R}^p. \quad (6.7)$$

Wir betrachten nun die Summe der Abweichungsquadrate

$$(\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta}) . \quad (6.8)$$

Es ergibt sich dann

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

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

$$(\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \geq 0 , \quad (6.9)$$

nimmt dieser Term genau dann seinen Minimalwert 0 an, wenn

$$(\hat{\beta} - \tilde{\beta}) = \mathbf{0}_p \Leftrightarrow \tilde{\beta} = \hat{\beta} . \quad (6.10)$$

Also gilt

$$\hat{\beta} = \arg \min_{\tilde{\beta}} (\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta}) . \quad (6.11)$$

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

$$\ell : \mathbb{R}^p \rightarrow \mathbb{R}, \tilde{\beta} \mapsto \ln p_{\tilde{\beta}}(\mathbf{y}) = \ln N(\mathbf{y}; X\tilde{\beta}, \sigma^2 I_n) , \quad (6.12)$$

wobei gilt, dass

$$\begin{aligned} \ln N(\mathbf{y}; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left( (2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left( -\frac{1}{2\sigma^2} (\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta}) \right) \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2} (\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta}) . \end{aligned} \quad (6.13)$$

Dabei hängt allein der Term  $-\frac{1}{2\sigma^2} (\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta})$  von  $\tilde{\beta}$  ab. Weil aber  $(\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta}) \geq 0$ , gilt aufgrund des negativen Vorzeichens, dass dieser Term maximal wird, wenn  $(\mathbf{y} - X\tilde{\beta})^T(\mathbf{y} - X\tilde{\beta})$  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

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}((X^T X)^{-1} X^T \mathbf{y}) = (X^T X)^{-1} X^T \mathbb{E}(\mathbf{y}) = (X^T X)^{-1} X^T X \beta = \beta . \quad (6.14)$$

□

Theorem 6.1 gibt mit

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y} \quad (6.15)$$

eine Formel an, um  $\beta$  anhand der Designmatrix  $X \in \mathbb{R}^{n \times p}$  und einer Realisierung des Datenvektors  $\mathbf{y} \in \mathbb{R}^n$  konkret zu schätzen. Als Zufallsvektor ist  $\hat{\beta}$  ein unverzerrter Schätzer von  $\beta$  und als Maximum-Likelihood-Schätzer insbesondere auch konsistent, asymptotisch normalverteilt und asymptotisch effizient. Wir werden an späterer Stelle sehen, dass  $\hat{\beta}$  sogar normalverteilt ist. Neben den genannten Eigenschaften hat  $\hat{\beta}$  noch weitere gute Eigenschaften. Zum Beispiel besitzt  $\hat{\beta}$  innerhalb der Klasse der linearen unverzerrten

Schätzer von  $\beta$  die kleinste Varianz. Diese Eigenschaft ist Kernaussage des *Gauss-Markov Theorems*, auf das wir hier aber nicht näher eingehen wollen.

Mithilfe des Betaparameterschätzers können wir die Begriffe der erklärten Daten, des Residuenvektors und der Residuen definieren, die wir an vielen Stellen benötigen werden.

**Definition 6.1** (Erklärte Daten, Residuenvektor und Residuen). Gegeben seien das ALM

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.16)$$

und der Betaparameterschätzer

$$\hat{\beta} := (X^T X)^{-1} X^T y. \quad (6.17)$$

Dann heißt der Zufallsvektor

$$\hat{y} := X\hat{\beta} \quad (6.18)$$

die *erklärten Daten*, der Zufallsvektor

$$\hat{\varepsilon} := y - \hat{y} \quad (6.19)$$

heißt *Residuenvektor* und für  $i = 1, \dots, n$  heißen die Komponenten dieses Zufallsvektors

$$\hat{\varepsilon}_i := y_i - \hat{y}_i \quad (6.20)$$

die *Residuen*.

•

Wir haben diese Begriffe bereits für den Fall einfacher linearer Regression mit Ausgleichsgerade eingeführt und generalisieren sie hier schlicht auf den Fall des Allgemeinen Linearen Modells. Erklärte Daten und Residuenvektor lassen sich als Matrixmultiplikationen mit dem Datenvektor schreiben, wie durch folgendes Theorem erklärt wird.

**Theorem 6.2** (Erklärte Daten und Residuenvektor als Matrixprodukte). Gegeben seien das ALM

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n). \quad (6.21)$$

sowie die erklärten Daten  $\hat{y}$  und der Residuenvektor  $\hat{\varepsilon}$ . Dann ergeben sich die erklärten Daten und der Residuenvektor als Matrixmultiplikationen des Datenvektors  $y$  von links

$$\begin{aligned} \hat{y} &= Py \\ \hat{\varepsilon} &= Ry, \end{aligned} \quad (6.22)$$

wobei die Projektionsmatrix  $P$  und die Residuen-bildende Matrix  $R$  wie folgt gegeben sind:

$$\begin{aligned} P &:= X(X^T X)^{-1} X^T \\ R &:= (I_n - P). \end{aligned} \quad (6.23)$$

◦

*Beweis.* Die erklärten Daten ergeben sich zu

$$\begin{aligned}\hat{y} &:= X\hat{\beta} \\ &= X(X^T X)^{-1} X^T y \\ &= Py.\end{aligned}\tag{6.24}$$

Der Residuenvektor ergibt sich zu

$$\begin{aligned}\hat{\varepsilon} &:= y - \hat{y} \\ &= y - X\hat{\beta} \\ &= y - X(X^T X)^{-1} X^T y \\ &= (I_n - X(X^T X)^{-1} X^T) y \\ &= Ry.\end{aligned}\tag{6.25}$$

□

## 6.2 Varianzparameterschätzung

Wir fassen die Frequentistische Punktschätzung des Varianzparameters in folgendem Theorem zusammen, das wir an dieser Stelle nicht beweisen wollen. Wir werden allerdings an späterer Stelle die Frequentistische Verteilung des Varianzparameterschätzers kennenlernen, aus der dieses Theorem folgt.

**Theorem 6.3** (Varianzparameterschätzer). *Gegeben sei das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n)\tag{6.26}$$

Dann ist

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}\tag{6.27}$$

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

◦

Theorem 6.3 gibt mit

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}\tag{6.28}$$

eine Formel an, um  $\sigma^2$  anhand der Designmatrix  $X \in \mathbb{R}^{n \times p}$ , des Betaparameterschätzers  $\hat{\beta} \in \mathbb{R}^p$  und einer Realisierung des Datenvektors  $y \in \mathbb{R}^n$  zu schätzen. Offenbar gilt mit Theorem 6.3, dass

$$\hat{\sigma}^2 = \frac{1}{n - p} \sum_{i=1}^n (y_i - (X\hat{\beta})_i)^2 = \frac{1}{n - p} \hat{\varepsilon}^T \hat{\varepsilon}.\tag{6.29}$$

$\hat{\sigma}^2$  wird also durch die Summe der quadrierten Residuen, also als eine Summe von Abweichungsquadraten geschätzt. Für einen Beweis von Theorem 6.3 verweisen wir zum Beispiel auf S. R. Searle (1971), S. R. Searle und Gruber (2017) oder Rencher und Schaalje (2008). Aus geometrischer Perspektive handelt es sich bei  $\hat{\sigma}^2$  um einen Kleinste-Quadrate-Schätzer (vgl. Christensen (2011)). Aus probabilistischer Perspektive handelt es sich bei  $\hat{\sigma}^2$  nicht um einen Maximum-Likelihood-Schätzer, sondern um einen Restricted-Maximum-Likelihood-Schätzer von  $\sigma^2$  (vgl. Harville (1977), Foulley (1993), Starke und Ostwald (2017)).

### 6.3 Unabhängig identisch normalverteilte Zufallsvariablen

Als erste Anwendung von Theorem 6.1 und Theorem 6.3 analysieren wir das Szenario von  $n$  unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter  $\mu \in \mathbb{R}$  und Varianzparameter  $\sigma^2$ :

$$y_i \sim N(\mu, \sigma^2) \text{ für } i = 1, \dots, n. \quad (6.30)$$

Schreibt man dieses Modell in seiner Designmatrixform, dann gilt, wie unten gezeigt,

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y} \text{ und } \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2. \quad (6.31)$$

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.

Für  $\hat{\beta}$  ergibt sich

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

Für  $\hat{\sigma}^2$  ergibt sich

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

Wir demonstrieren die Parameterschätzung in diesem Szenario in folgendem **R**-Code.

```

# Modellformulierung
library(MASS) # multivariate Normalverteilung
n = 12 # Anzahl Datenpunkte
p = 1 # Anzahl Betaparameter
X = matrix(rep(1,n), nrow = n) # n x p 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 des n-dimensionalen ZVs y

# 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

```

```

beta : 2
hat{beta} : 2.066348
sigsqr : 1
hat{sigsqr}: 0.9944977

```

Die Frequentistische Bedeutung der Schätzerunverzerrtheit in diesem Szenario simuliert folgender **R**-Code.

```

# Modellformulierung
library(MASS) # multivariate Normalverteilung
n = 12 # Anzahl Datenpunkte
p = 1 # Anzahl Betaparameter
X = matrix(rep(1,n), nrow = n) # n x p 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 Datenrealisierungen
beta_hat = rep(NA,n,nsim) # \hat{\beta} Realisierungsarray
sigsqr_hat = rep(NA,n,nsim) # \hat{sigsqr} Realisierungsarray
for(i in 1:nsim){ # Simulationsiterationen
  y = mvrnorm(1, X %>% beta, sigsqr*I_n) # Datenrealisierung
  beta_hat[i] = solve(t(X) %>% X) %>% t(X) %>% y # Betaparameterschätzer
  eps_hat = y - X %>% beta_hat[i] # Residuenvektor
  sigsqr_hat[i] = (t(eps_hat) %>% eps_hat) / (n-p) # Varianzparameterschätzer
}

```

```

wahrer, aber unbekannter, Betaparameter : 2
geschätzter Erwartungswert des Betaparameterschätzers : 1.999488
wahrer, aber unbekannter, Varianzparameter : 1
geschätzter Erwartungswert des Varianzparameterschätzers : 1.004968

```

## 6.4 Einfache lineare Regression

Als zweite Anwendung von Theorem 6.1 und Theorem 6.3 analysieren wir das Szenario der einfachen linearen Regression:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n. \quad (6.32)$$

Basierend auf der Designmatrixform Gleichung 5.15 dieses Modells ergibt sich, wie unten gezeigt,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix} \text{ und } \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2, \quad (6.33)$$

wobei  $\bar{x}$  und  $\bar{y}$  die Stichprobenmittel der  $x_1, \dots, x_n$  und  $y_1, \dots, y_n$ ,  $c_{xy}$  die Stichprobenkovarianz der  $x_1, \dots, x_n$  und  $y_1, \dots, y_n$  und  $s_x^2$  die Stichprobenvarianz der  $x_1, \dots, x_n$  bezeichnen. Wie in Kapitel 1 sind die Bezeichnungen Stichprobenkovarianz und Stichprobenvarianz

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.

Wir halten also fest, dass für eine parametrische Designmatrixspalte sich der entsprechende Betaparameterschätzer aus der Stichprobenkovarianz der zugehörigen Spalte mit den Daten, geteilt durch die Stichprobenvarianz der entsprechenden Spalte, ergibt und somit einer standardisierten Stichprobenkovarianz entspricht. Ein Vergleich mit den Parametern der Ausgleichsgerade in Kapitel 1 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 bei gegebenem Datenvektor  $y$  den Wert

$$q(\tilde{\beta}) = \sum_{i=1}^n (y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2 = (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (6.34)$$

hinsichtlich  $\tilde{\beta}$  minimieren.

Um die Form des Betaparameterschätzers in Gleichung 6.33 herzuleiten, halten wir zunächst fest, dass

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \bar{y} - \sum_{i=1}^n \bar{x} y_i + \sum_{i=1}^n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \bar{y} n \bar{x} - \bar{x} n \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}, \end{aligned} \quad (6.35)$$

Weiterhin halten wir fest, dass

$$\begin{aligned}
\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2x_i\bar{x} + \sum_{i=1}^n \bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - n\bar{x}^2.
\end{aligned} \tag{6.36}$$

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

$$\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T y \\
&= \left( \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \left( \begin{matrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{matrix} \right)^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\
&= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.
\end{aligned} \tag{6.37}$$

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

$$\frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \tag{6.38}$$

weil

$$\begin{aligned}
&\frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \\
&= \frac{1}{s_x^2} \begin{pmatrix} \frac{ns_x^2}{n} + n\bar{x}^2 - n\bar{x}^2 & \frac{s_x^2 n\bar{x}}{n} + n\bar{x}^2\bar{x} - \bar{x} \sum_{i=1}^n x_i^2 \\ -\bar{x}n + n\bar{x} & -n\bar{x}^2 + \sum_{i=1}^n x_i^2 \end{pmatrix} \\
&= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & s_x^2\bar{x} - \bar{x} (\sum_{i=1}^n x_i^2 - n\bar{x}^2) \\ 0 & \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{pmatrix} \\
&= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & s_x^2\bar{x} - \bar{x}s_x^2 \\ 0 & s_x^2 \end{pmatrix} \\
&= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & 0 \\ 0 & s_x^2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned} \tag{6.39}$$

Es ergibt sich also

$$\begin{aligned}
 \hat{\beta} &= \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_x^2} & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right) n\bar{y} - \frac{\bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i}{s_x^2} - \frac{n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{n\bar{y}}{n} + \frac{\bar{x}^2 n\bar{y}}{s_x^2} - \frac{\bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
 &= \begin{pmatrix} \bar{y} + \frac{\bar{x} n \bar{x} \bar{y} - \bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
 &= \begin{pmatrix} \bar{y} - \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \bar{x} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\
 &= \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix}.
 \end{aligned} \tag{6.40}$$

Wir demonstrieren die Parameterschätzung in diesem Szenario in folgendem **R**-Code. Man beachte die weitgehende Übereinstimmung mit der Implementation der Parameterschätzung im Szenario der unabhängig und identisch normalverteilten Zufallsvariablen – lediglich die Designmatrix und die Dimension des Betaparameters ändern sich.

```

# Modellformulierung
library(MASS)
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

# multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Prädiktorwerte
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n)

# Parameterschätzung
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y
eps_hat = y - X %*% beta_hat
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)

beta      : 0 1
hat{beta} : 1.07092 0.8417131
sigsqr    : 1
hat{sigsqr}: 0.5936961

```

Analog zum Szenario der unabhängig und identisch normalverteilten Zufallsvariablen kann auch hier die Frequentistische Bedeutung der Schätzerunverzerrtheit simuliert werden.

```

# Modellformulierung
library(MASS)
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

# multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Prädiktorwerte
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Frequentistische Simulation
nsim = 1e4
beta_hat = matrix(rep(NA,n*p*nsim), nrow = p)
sigsqr_hat = rep(NA,nsim)
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)
}

```

```
wahrer, aber unbekannter, Betaparameter           : 0 1
geschätzter Erwartungswert des Betaparameterschätzers : 0.001888341 0.9998024
wahrer, aber unbekannter, Varianzparameter         : 1
geschätzter Erwartungswert des Varianzparameterschätzers : 0.9964691
```

## 6.5 Frequentistische Schätzerverteilungen

Wir dokumentieren die Frequentistische Verteilung des Betaparameterschätzers in folgendem Theorem.

**Theorem 6.4** (Frequentistische Verteilung des Betaparameterschätzers). *Gegeben seien das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.41)$$

und der Betaparameterschätzer

$$\hat{\beta} := (X^T X)^{-1} X^T y \quad (6.42)$$

Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (6.43)$$

◦

*Beweis.* Die Aussage des Theorems folgt direkt mit dem Theorem zur linear-affinen Transformation von multivariaten Normalverteilungen aus Kapitel 4. Speziell gilt hier:

$$\hat{\beta} \sim N((X^T X)^{-1} X^T X \beta, (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T). \quad (6.44)$$

Der Erwartungswertparameter vereinfacht sich dann zu

$$(X^T X)^{-1} X^T X \beta = \beta. \quad (6.45)$$

Der Kovarianzmatrixparameter vereinfacht sich wie folgt:

$$\begin{aligned} (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T &= (X^T X)^{-1} X^T (\sigma^2 I_n) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned} \quad (6.46)$$

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

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (6.47)$$

□

Mit Theorem 6.4 folgt insbesondere auch für den Erwartungswert und die Kovarianzmatrix des Betaparameterschätzers, dass

$$\mathbb{E}(\hat{\beta}) = \beta \text{ und } \mathbb{C}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \quad (6.48)$$

Als Diagonalelemente von  $\mathbb{C}(\hat{\beta})$  hängen die Varianzen der Betaparameterschätzerkomponenten also sowohl vom Varianzparameter der Fehlervariablen als auch von der Designmatrix ab. Insbesondere bei festem wahren, aber unbekanntem  $\sigma^2 > 0$  kann also die

Designmatrix so gewählt werden, dass die Varianz der Betaparameterschätzerkomponenten minimiert wird.

Die Frequentistische Verteilung des Varianzparameterschätzers dokumentieren wir in folgendem Theorem.

**Theorem 6.5** (Frequentistische Verteilung des Varianzparameterschätzers). *Gegeben seien das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.49)$$

und der Variantparameterschätzer

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^\top (y - X\hat{\beta})}{n - p}. \quad (6.50)$$

Dann gilt

$$\frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n - p). \quad (6.51)$$

◦

*Beweis.* Wir zeigen zunächst, dass sich  $(n - p)\hat{\sigma}^2$  als quadratische Form von  $y$  schreiben lässt:

$$\begin{aligned} (n - p)\hat{\sigma}^2 &= (n - p) \frac{(y - X\hat{\beta})^\top (y - X\hat{\beta})}{n - p} = (y - X\hat{\beta})^\top (y - X\hat{\beta}) \\ &= (y - \hat{y})^\top (y - \hat{y}) = \hat{\varepsilon}^\top \hat{\varepsilon} = (Ry)^\top (Ry) = y^\top R^\top Ry \\ &\text{mit } R := I_n - P = I_n - X(X^\top X)^{-1} X^\top. \end{aligned} \quad (6.52)$$

Im nächsten Schritt bemerken wir, dass die Residuen-bildende  $R$  symmetrisch ist

$$\begin{aligned} R^\top &= (I_n - P)^\top = (I_n - X(X^\top X)^{-1} X^\top)^\top = I_n^\top - (X(X^\top X)^{-1} X^\top)^\top \\ &= I_n - X(X^\top X)^{-1} X^\top = I_n - P = R \end{aligned} \quad (6.53)$$

und dass  $R$  idempotent ist, d.h. mit sich selbst multipliziert wieder  $R$  ergibt

$$\begin{aligned} RR &= (I_n - P)(I_n - P) = I_n - P - P + PP \\ &= I_n - 2P + X(X^\top X)^{-1} X^\top X(X^\top X)^{-1} X^\top \\ &= I_n - 2P + X(X^\top X)^{-1} X^\top = I_n - 2P + P = I_n - P = R, \end{aligned} \quad (6.54)$$

sodass gilt:

$$(n - p)\hat{\sigma}^2 = y^\top R^\top Ry = y^\top RRy = y^\top Ry. \quad (6.55)$$

Nun sind die Anwendungsbedingungen von Cochrans Theorem für multivariate Normalverteilungen gegeben:

$$\xi \sim N(\mu, \sigma^2 I_n) \quad \Rightarrow \quad y := \xi^\top A \xi / \sigma^2 \sim \chi^2(\mu^\top A \mu, \text{rg}(A)). \quad (6.56)$$

Im vorliegenden Fall haben wir:

$$y \sim N(X\beta, \sigma^2 I_n) \quad \text{und} \quad \frac{n - p}{\sigma^2} \hat{\sigma}^2 = y^\top Ry / \sigma^2. \quad (6.57)$$

Also gilt:

$$\frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2((X\beta)^\top R(X\beta), \text{rg}(R)). \quad (6.58)$$

Für den Nichtzentralitätsparameter ergibt sich:

$$\begin{aligned}
(X\beta)^T R(X\beta) &= \beta^T X^T (I_n - P) X \beta \\
&= \beta^T X^T (I_n - X(X^T X)^{-1} X^T) X \beta \\
&= \beta^T (X^T X - X^T X (X^T X)^{-1} X^T X) \beta \\
&= \beta^T (X^T X - X^T X) \beta \\
&= \beta^T \mathbf{0}_{pp} \beta \\
&= 0.
\end{aligned} \tag{6.59}$$

Da  $R$  idempotent ist, ist ihr Rang gleich ihrer Spur und für den Freiheitsgradparameter ergibt sich:

$$\begin{aligned}
\text{rg}(R) = \text{sp}(R) &= \text{sp}(I_n - P) \\
&= \text{sp}(I_n - X(X^T X)^{-1} X^T) \\
&= \text{sp}(I_n) - \text{sp}(X(X^T X)^{-1} X^T) \\
&= \text{sp}(I_n) - \text{sp}(X^T X (X^T X)^{-1}) \\
&= \text{sp}(I_n) - \text{sp}(I_p) \\
&= n - p.
\end{aligned} \tag{6.60}$$

Zusammengenommen folgt daraus:

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(0, n-p). \tag{6.61}$$

Da eine nichtzentrale Chi-Quadrat-Verteilung mit Nichtzentralitätsparameter 0 äquivalent zu einer Chi-Quadrat-Verteilung mit gleichem Freiheitsgradparameter ist, gilt schließlich:

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p). \tag{6.62}$$

□

Da es sich bei  $(n-p)\hat{\sigma}^2$  um eine Summe normalverteilter Zufallsvariablen handelt, liegt die  $\chi^2$ -Verteilung im Lichte der  $\chi^2$ -Transformation bei normalverteilten Zufallsvariablen (vgl. *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*) bereits nahe. Allerdings ist  $\hat{\sigma}^2$  selbst nicht  $\chi^2$  verteilt, sondern lediglich seine durch Multiplikation mit  $\frac{n-p}{\sigma^2}$  skalierte Version. Wir wollen die Frequentistischen Schätzerverteilungen aus Theorem 6.4 und Theorem 6.5 noch an den beiden Standardbeispielen verdeutlichen.

### Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^n, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \tag{6.63}$$

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. Theorem 6.4 impliziert dann mit

$$(X^T X)^{-1} = (\mathbf{1}_n^T \mathbf{1}_n)^{-1} = \frac{1}{n}, \tag{6.64}$$

dass

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \tag{6.65}$$

Das Stichprobenmittel von  $n$  unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter  $\mu$  und Varianzparameter  $\sigma^2$  ist also normalverteilt

mit Erwartungswertparameter  $\mu$  und Varianzparameter  $\sigma^2/n$ . Wir haben diese Tatsache bereits im Kontext der Transformationen der Normalverteilungen unter dem Begriff der *Mittelwerttransformation* kennengelernt (vgl. *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*).

## Beispiel (2) Einfache lineare Regression

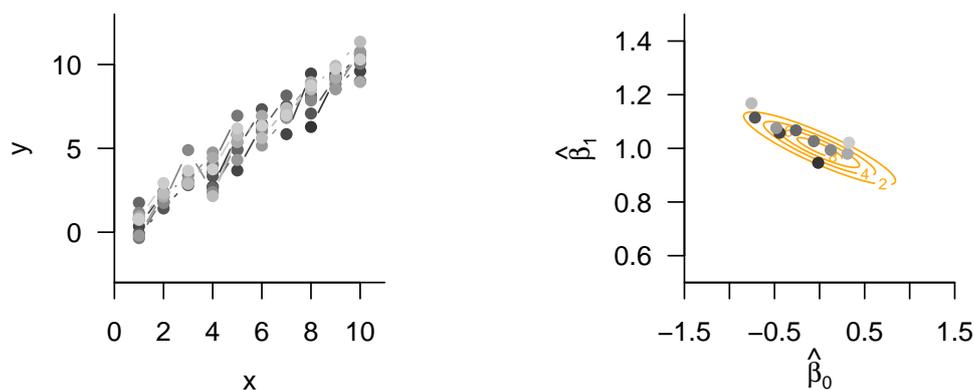
Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0 \quad (6.66)$$

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

$$\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \text{ mit } s_x^2 := \sum_{i=1}^n (x_i - \bar{x})^2. \quad (6.67)$$

Die Varianz des Offsetparameterschätzers hängt damit sowohl von der Summe der quadrierten Differenzen der Werte der unabhängigen Variable von ihrem Stichprobenmittel und dem Stichprobenmittel der Werte der unabhängigen Variable selbst ab. Die Varianz des Steigungsparameterschätzers hängt dagegen nur von der Summe der quadrierten Differenzen der unabhängigen Variable von ihrem Stichprobenmittel ab. Die Kovarianz von Offset- und Steigungsparameterschätzern schließlich hängt vom Mittelwert der Werte der unabhängigen Variable ab.



**Abbildung 6.1** Frequentistische Betaparameterschätzerverteilung bei einfacher linearer Regression.

Folgender **R**-Code simuliert die frequentistischen Verteilungen von Beta- und Varianzparameterschätzern im Szenario der einfachen linearen Regression.

```
# Modellformulierung
library(MASS)
n = 10
p = 2
x = 1:n
X = matrix(c(rep(1,n),x), nrow = n)

# multivariate Normalverteilung
# Anzahl von Datenpunkten
# Anzahl von Betparametern
# Prädiktorwerte
# n x p Designmatrix
```

```

I_n      = diag(n)                # n x n Einheitsmatrix
beta     = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter Betaparameter
sigsqr   = 0.5                   # 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] = solve(t(X) %*% X) %*% t(X) %*% y[,i] # \hat{\beta} = (X^T X)^{-1} X^T y
}

```

Abbildung 6.1 A zeigt 10 Realisationen des Modells einer einfachen linearen Regression und Abbildung 6.1 B zeigt die entsprechenden Betaparameterschätzerrealisationen sowie die analytische Verteilung des Betaparameterschätzers.

## 6.6 Literaturhinweise

Plackett (1949) gibt einen historischen Überblick zur Entwicklung der Betaparameterschätzung und insbesondere des Gauss-Markov-Theorems. Das Problem der Varianzparameterschätzung im Rahmen des ALMs im Sinne der Restricted-Maximum-Likelihood-Methode erscheint zunächst in Patterson und Thompson (1971) (vgl. Harville (1977), Verbyla (1990)) und bleibt, in verallgemeinerten ALMs, Gegenstand aktueller Forschung (vgl. Lindholm und Wahl (2020)).

## 6.7 Selbstkontrollfragen

1. Geben Sie das Theorem zum Betaparameterschätzer wieder.
2. Warum ist der Betaparameterschätzer ein Maximum-Likelihood-Schätzer?
3. Geben Sie das Theorem zum Varianzparameterschätzer wieder.
4. Warum ist der Varianzparameterschätzer kein Maximum-Likelihood-Schätzer?
5. Geben Sie die Definition der erklärten Daten und des Residuenvektors wieder.
6. Geben Sie das Theorem zu erklärten Daten und Residuenvektor als Matrixprodukte wieder.
7. Geben Sie die Parameterschätzer bei unabhängigen und identisch normalverteilten Zufallsvariablen an.
8. Geben Sie die Parameterschätzer bei einfacher linearer Regression an.
9. Geben Sie die Definition einer  $\chi^2$ -Zufallsvariable wieder.
10. Geben Sie die Definition einer nichtzentralen  $\chi^2$ -Zufallsvariable wieder.
11. Geben Sie das Theorem zur Verteilung des Betaparameterschätzers wieder.
12. Geben Sie das Theorem zur Verteilung des Varianzparameterschätzers wieder.

## 7 T-Statistiken

In diesem Abschnitt führen wir T-Statistiken als Maße zur Evaluation von Betaparameterschätzern im ALM ein. T-Statistiken setzen dabei die geschätzten Effekte des Betaparameterschätzers in Relation zur durch den Varianzparameterschätzer geschätzten Residualvariabilität. Der Wert einer T-Statistik ist also zunächst einmal einfach als Signal-zu-Rauschen-Verhältnis (*signal-to-noise ratio*) zu verstehen.

T-Statistiken erlauben weiterhin die Evaluation von Linearkombinationen der Komponenten des Betaparameterschätzers im Sinne Frequentistischer Konfidenzintervalle und Hypothesentests. Wir betrachten hier zunächst nur die funktionale Form von T-Statistiken und ihre Frequentistische Verteilung zum Zwecke der Konfidenzintervallbestimmung. Der Einsatz von T-Teststatistiken im Rahmen von Einstichproben- und Zweistichproben-T-Tests ist dann das Thema von Kapitel 9.

### 7.1 Definition und Beispiele

Vor dem Hintergrund des ALMs und seiner Parameterpunktschätzer definieren wir die T-Statistik wie folgt.

**Definition 7.1** (T-Statistik). Gegeben seien das ALM

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7.1)$$

sowie die Betaparameter- und Varianzparameterschätzer

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}. \quad (7.2)$$

Dann ist für einen Kontrastgewichtsvektor  $c \in \mathbb{R}^p$  und einen Nullparameter  $\beta_0 \in \mathbb{R}^p$  die *T-Statistik* definiert als

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \quad (7.3)$$

•

Geeignete Wahlen des Kontrastgewichtvektors  $c \in \mathbb{R}^p$  und des Parameters  $\beta_0 \in \mathbb{R}^p$  erlauben eine Vielzahl von Einsatzmöglichkeiten der T-Statistik. Betrachten wir zunächst den Kontrastgewichtsvektor. Offenbar dient der Kontrastgewichtsvektor dazu, den Zufallsvektor  $\hat{\beta} \in \mathbb{R}^p$  in die Zufallsvariable  $c^T \hat{\beta} \in \mathbb{R}$  zu transformieren und stellt damit sicher, dass es sich bei der T-Statistik um einen Skalar handelt. Weiterhin erlaubt die Wahl von  $p$ -dimensionalen Einheitsvektoren für den Kontrastgewichtsvektor die Auswahl einzelner Komponenten des Betaparameters zur Evaluation mithilfe der T-Statistik. Schließlich

erlaubt eine generelle Wahl des Kontrastgewichtsvektors die Evaluation beliebiger Linearkombination der Betaparameterkomponenten, wie zum Beispiel Differenzen einzelner Komponenten.

Beispielhaft seien für  $\hat{\beta} \in \mathbb{R}^2$  hier folgende Möglichkeiten für die Wahl von  $c \in \mathbb{R}^2$  hinsichtlich des Skalarproduktes  $c^T \hat{\beta}$  aufgeführt:

$$c := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow c^T \hat{\beta} = \hat{\beta}_1, \quad c := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow c^T \hat{\beta} = \hat{\beta}_2, \quad c := \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow c^T \hat{\beta} = \hat{\beta}_1 - \hat{\beta}_2. \quad (7.4)$$

Die Wahl des Parameters  $\beta_0 \in \mathbb{R}^p$  eröffnet die Möglichkeit, die T-Statistik unterschiedlich einzusetzen. Wählt man zum Beispiel  $\beta_0 := 0_p$ , so erhält man mit der T-Statistik eine Deskriptivstatistik, die es erlaubt, geschätzte Regressoreffekte, also Komponenten oder Linearkombinationen von  $\hat{\beta}$ , im Sinne eines Signal-zu-Rauschen-Verhältnisses in Bezug zu der durch  $\hat{\sigma}^2$  quantifizierten Residualdatenvariabilität zu setzen. Der Nenner der T-Statistik stellt dabei sicher, dass insbesondere die adäquate (Ko-)Standardabweichung der entsprechenden Betaparameterkomponentenkombination als Bezugsgröße dient, da es sich bei  $\hat{\sigma}^2 (X^T X)^{-1}$  bekanntlich um die Kovarianz des Betaparameterschätzers handelt (siehe Theorem 6.4).

Wählt man für  $\beta_0$  dagegen  $\beta$ , also den wahren, aber unbekanntem, Betaparameterwert, so eröffnet die T-Statistik die Möglichkeit, für einzelne Komponenten des Betaparametervektors Konfidenzintervalle zu bestimmen. Wir vertiefen diesen Aspekt der T-Statistik in Kapitel 7.2. Deklariert man schließlich  $\beta_0$  im Kontext eines Testszenarios als das Element einer Nullhypothese, so eröffnet die T-Statistik die hypothesentestbasierte Inferenz über Betaparameterkomponenten und ihrer Linearkombinationen. Anwendungsfälle dieser Art diskutieren wir ausführlich in Kapitel 9.

Die Anwendung der T-Statistik zum Zwecke der Frequentistischen Inferenz im Sinne von Konfidenzintervallen und Hypothesentests basiert dabei natürlich auf der Frequentistischen Verteilung der T-Statistik vor dem Hintergrund des ALMs, die Inhalt des folgenden Theorems ist.

**Theorem 7.1** (Verteilung der T-Statistik). *Gegeben seien das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7.5)$$

*sowie die Betaparameter- und Varianzparameterschätzer*

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}. \quad (7.6)$$

*Schließlich sei für einen Kontrastgewichtsvektor  $c \in \mathbb{R}^p$  und einen Nullparameter  $\beta_0 \in \mathbb{R}^p$  die T-Statistik*

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \quad (7.7)$$

*Dann gilt:*

$$T \sim t(\delta, n - p) \text{ mit } \delta = \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}}. \quad (7.8)$$

◦

*Beweis.* Wir wissen, dass der Betaparameterschätzer einer multivariaten Normalverteilung folgt (siehe Theorem 6.4):

$$\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1}). \quad (7.9)$$

Mit dem Theorem zur linear-affinen Transformation multivariat normalverteilter Zufallsvektoren (siehe Theorem 4.3) gilt daher

$$c^T \hat{\beta} - c^T \beta_0 \sim N(c^T \beta - c^T \beta_0, \sigma^2 c^T (X^T X)^{-1} c) \quad (7.10)$$

sowie

$$Z = \frac{1}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} (c^T \hat{\beta} - c^T \beta_0) \sim N\left(\frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}}, 1\right). \quad (7.11)$$

Wir wissen darüber hinaus, dass folgende Funktion des Varianzparameterschätzers einer Chi-Quadrat-Verteilung folgt (siehe Theorem 6.5):

$$U = \frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p). \quad (7.12)$$

Definitionsgemäß ist der Quotient aus der Zufallsvariable  $Z$  und der Wurzel der Zufallsvariable  $U$ , geteilt durch den Freiheitsgradparameter, eine nichtzentrale  $t$ -Zufallsvariable

$$T = \frac{Z}{\sqrt{U/(n-p)}} \sim t(\delta, n-p), \quad (7.13)$$

wobei der Nichtzentralitätsparameter der nichtzentralen Chi-Quadrat-Verteilung von  $T$  durch den Erwartungswertparameter der Normalverteilung von  $Z$  gegeben ist

$$\delta = \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} \quad (7.14)$$

und sich die T-Statistik wie folgt ergibt:

$$\begin{aligned} T &= \frac{\frac{1}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} (c^T \hat{\beta} - c^T \beta_0)}{\sqrt{\frac{n-p}{\sigma^2} \hat{\sigma}^2 / (n-p)}} \\ &= \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c} \sqrt{\hat{\sigma}^2 / \sigma^2}} \\ &= \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \end{aligned} \quad (7.15)$$

□

Im Allgemeinen ist die T-Statistik also nichtzentral  $t$ -verteilt. Gilt dabei, wie bei der Bestimmung von Konfidenzintervallen (vgl. Kapitel 7.2)  $\beta_0 := \beta$  oder gilt in einem Test-szenario bei Zutreffen der Nullhypothese  $\beta := \beta_0$  (vgl. Kapitel 9), so ist die T-Statistik sogar  $t$ -verteilt, jeweils mit Freiheitsgradparameter  $n-p$ .

Gilt in einem Test-szenario dagegen, dass die Nullhypothese nicht zutrifft, so kann die Verteilung der T-Statistik aus Theorem 7.1 zur Herleitung der Testgütefunktion und damit zur Bestimmung der Power eines T-Tests genutzt werden (vgl. Kapitel 9). Wir werden diese Aspekte an gegebener Stelle vertiefen. An dieser Stelle wollen wir die T-Statistik und ihre Verteilung zunächst nur an den Beispielen der unabhängigen identisch normalverteilten Zufallsvariablen und der einfachen linearen Regression illustrieren.

### Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

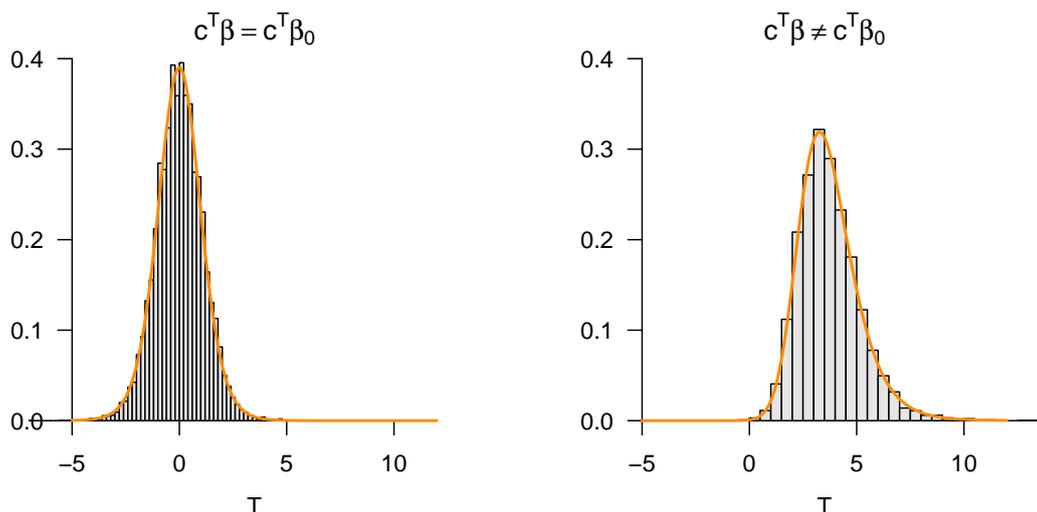
Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0 \quad (7.16)$$

das ALM-Szenario unabhängiger und identisch normalverteilter Zufallsvariablen und es seien  $c := 1$  und  $\beta_0 := \mu_0 \in \mathbb{R}$ . Dann gilt für die T-Statistik:

$$T = \frac{c^T \hat{\beta} - c^T \mu_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{\mathbf{1}^T \bar{v} - \mathbf{1}^T \mu_0}{\sqrt{s_v^2 \mathbf{1}^T (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}}} = \sqrt{n} \left( \frac{\bar{v} - \mu_0}{s_v} \right). \quad (7.17)$$

Dies entspricht offenbar der bekannten Statistik des Einstichproben-T-Tests (vgl. *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*). Wie diese nimmt die hier betrachtete T-Statistik, bei Konstanz der jeweils komplementären Terme, große absolute Werte für eine große absolute Differenz von  $\bar{v} - \mu_0$  (oft als Effekt bezeichnet) sowie für kleine Werte von  $s_v^2$  (also eine geringe Datenvariabilität) und einen großen Wert von  $n$  (also einen großen Stichprobenumfang) an.



**Abbildung 7.1** Verteilungen der T-Statistik bei unabhängig und identisch normalverteilten Zufallsvariablen.

Folgender **R**-Code simuliert die Frequentistische Verteilung dieser T-Statistik für die Fälle  $\beta = \beta_0$  und  $\beta \neq \beta_0$ .

```
# Modellformulierung
library(MASS)
n = 12
p = 1
X = matrix(c(rep(1,n)), nrow = n)
I_n = diag(n)
beta = c(0,1)
nscn = length(beta)
sigsqr = 1
c = 1
beta_0 = 0

# multivariate Normalverteilung
# Anzahl von Datenpunkten
# Anzahl von Betaparametern
# n x p Designmatrix
# n x n Einheitsmatrix
# wahre, aber unbekannte Betaparameter
# Anzahl wahrer, aber unbekannter Hypothesenszenarien
# wahrer, aber unbekannter Varianzparameter
# Kontrastvektor von Interesse
# Nullhypothesenbetaparameter

# Frequentistische Simulation
nsim = 1e4
delta = rep(NaN, nscn)
Tee = matrix(rep(NaN, nscn*nsim), ncol = nscn)
for(s in 1:nscn){
```

```

delta[s] = ((t(c) %>% beta[s] - t(c) %>% beta_0)/
  sqrt(sigsqr*t(c)%solve(t(X)%X)%c)) # Nichtzentralitätsparameter
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))
}
}

```

Abbildung 7.1 A und B zeigen die resultierenden simulierten und analytischen Verteilungen der T-Statistik.

## Beispiel (2) Einfache lineare Regression

In diesem Beispiel wollen wir nicht auf die spezifische Form der T-Statistik eingehen, aber anhand einer Simulation demonstrieren, wie sich das Prinzip der T-Statistik im Kontext der einfachen linearen Regression darstellt. Dazu betrachten wir das bekannte Beispielmodell der einfachen linearen Regression (vgl. Kapitel 5), in diesem Fall mit den wahren, aber unbekannt Parameterwerten  $\beta_A := (1, 0)$  und  $\beta_B := (1, 1)$ . Weiterhin betrachten wir den Kontrastgewichtsvektor  $c := (0, 1)$ , so dass die T-Statistik zu Evaluation des Steigungsparameters der einfachen linearen Regression genutzt werden kann. Schließlich betrachten wir in beiden Fällen den Parameter  $\beta_0 := (0, 0)^T$ , so dass im Fall von  $\beta_A$  gilt, dass  $c^T \beta = c^T \beta_0$  und im Fall von  $\beta_B$  gilt, dass  $c^T \beta \neq c^T \beta_0$ .

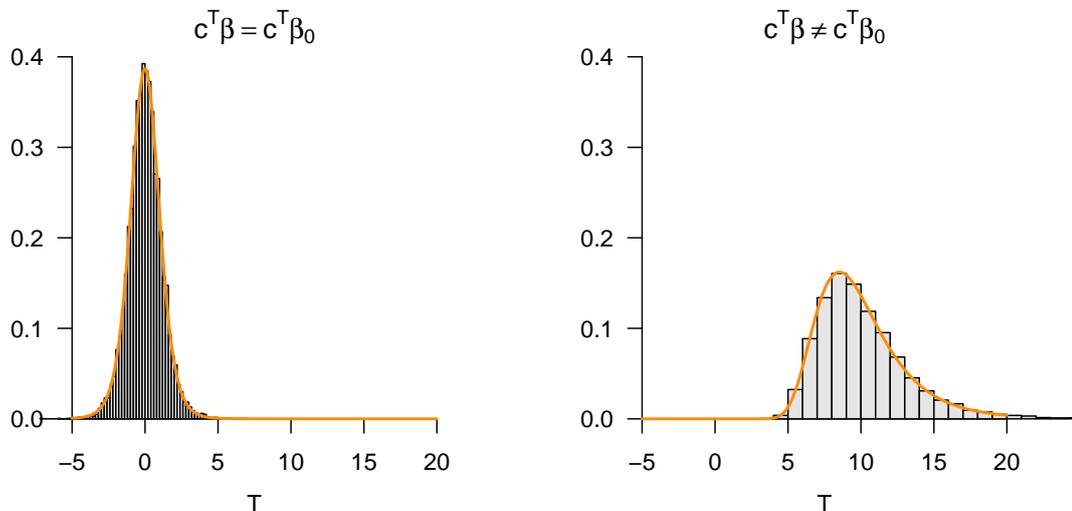


Abbildung 7.2 Verteilungen der T-Statistik bei einfacher linearer Regression.

Folgender **R**-Code implementiert die skizzierten Szenarien.

```

# Modellformulierung
library(MASS) # multivariate Normalverteilung
n = 10 # Anzahl von Datenpunkten
p = 2 # Anzahl von Betaparametern
x = 1:n # Prädiktorwerte
X = matrix(c(rep(1,n),x), ncol = p) # n x p Designmatrix
I_n = diag(n) # n x 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(NA, nscn)                       # Anzahl Nichtzentralitätsparameter
Tee       = matrix(rep(NA, nscn*nsim), ncol = nscn) # T-Teststatistik-Realisierungsarray
for(s in 1:nscn){
  delta[s] = ((t(c) %>% beta[,s] - t(c) %>% beta_0)/
             sqrt(sigsqr*t(c) %>% solve(t(X) %>% X) %>% c))
             # Nichtzentralitätsparameter
  for(i in 1:nsim){
    y       = mvrnorm(1, X %>% beta[,s], sigsqr*I_n) # Simulationsiterationen
    beta_hat = solve(t(X) %>% X) %>% t(X) %>% y      # y
    eps_hat  = y - X %>% beta_hat                    # \hat{\beta}
    sigsqr_hat = (t(eps_hat) %>% eps_hat)/(n-p)      # \hat{\epsilon}
    Tee[i,s] = ((t(c) %>% beta_hat - t(c) %>% beta_0)/
               sqrt(sigsqr_hat*t(c) %>% solve(t(X) %>% X) %>% c)) # \hat{\sigma}^2
  }
}

```

Abbildung 7.2 A und B zeigen die resultierenden simulierten und analytischen Verteilungen der T-Statistik.

## 7.2 Konfidenzintervalle für Betaparameterkomponenten

Mithilfe der T-Statistik können Konfidenzintervalle für die Komponenten des Betaparametervektors bestimmt werden. Das folgende Theorem ist die zentrale Aussage dieses Abschnitts.

**Theorem 7.2** (Konfidenzintervalle für Betaparameterkomponenten). *Gegeben seien das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7.18)$$

sowie der Betaparameterschätzer  $\hat{\beta}$  und der Varianzparameterschätzer  $\hat{\sigma}^2$ . Für ein  $\gamma \in ]0, 1[$  sei

$$t_\gamma := \Psi^{-1} \left( \frac{1 + \gamma}{2}; n - p \right). \quad (7.19)$$

Schließlich sei  $\lambda_j$  das  $j$ te Diagonalelement von  $(X^T X)^{-1}$  für  $j = 1, \dots, p$ :

$$\lambda_j := \left( (X^T X)^{-1} \right)_{jj}. \quad (7.20)$$

Dann ist

$$\kappa_j := \left[ \hat{\beta}_j - \hat{\sigma} \sqrt{\lambda_j} t_\gamma, \hat{\beta}_j + \hat{\sigma} \sqrt{\lambda_j} t_\gamma \right] \quad (7.21)$$

ein  $\gamma$ -Konfidenzintervall für die  $j$ te Komponente  $\beta_j$  des Betaparameters  $\beta = (\beta_1, \dots, \beta_p)^T$ , für  $j = 1, \dots, p$ .

◦

*Beweis.* Wir müssen zeigen, dass

$$\mathbb{P}(\kappa_j \ni \beta_j) = \gamma. \quad (7.22)$$

Dazu halten wir zunächst fest, dass für alle  $j = 1, \dots, p$  bei Wahl von  $\beta_0 = \beta$  und  $c := e_j$  nach dem Theorem zur T-Statistik für  $T \sim t(\delta, n - p)$  gilt, dass

$$T = \frac{e_j^T \hat{\beta} - e_j^T \beta}{\sqrt{\hat{\sigma}^2 e_j^T (X^T X)^{-1} e_j}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \left( (X^T X)^{-1} \right)_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{\lambda_j}} =: T_j. \quad (7.23)$$

und

$$\delta = \frac{e_j^T \beta - e_j^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 e_j^T (X^T X)^{-1} e_j}} = 0 \quad (7.24)$$

Damit gilt dann auch sofort, dass  $T_j \sim t(n-p)$ . Weiterhin erinnern wir daran (vgl. *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*), dass per Definition von  $t_\gamma$  gilt, dass

$$\mathbb{P}(-t_\gamma \leq T_j \leq t_\gamma) = \gamma. \quad (7.25)$$

Aus der Definition eines  $\gamma$ -Konfidenzintervalls folgt dann

$$\begin{aligned} \gamma &= \mathbb{P}(-t_\gamma \leq T_j \leq t_\gamma) \\ &= \mathbb{P}\left(-t_\gamma \leq \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{\lambda_j}} \leq t_\gamma\right) \\ &= \mathbb{P}\left(-t_\gamma \hat{\sigma} \sqrt{\lambda_j} \leq \hat{\beta}_j - \beta_j \leq t_\gamma \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left(-\hat{\beta}_j - t_\gamma \hat{\sigma} \sqrt{\lambda_j} \leq -\beta_j \leq -\hat{\beta}_j + t_\gamma \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left(\hat{\beta}_j + t_\gamma \hat{\sigma} \sqrt{\lambda_j} \geq \beta_j \geq \hat{\beta}_j - t_\gamma \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left(\hat{\beta}_j - t_\gamma \hat{\sigma} \sqrt{\lambda_j} \leq \beta_j \leq \hat{\beta}_j + t_\gamma \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left([\hat{\beta}_j - \hat{\sigma} \sqrt{\lambda_j} t_\gamma, \hat{\beta}_j + \hat{\sigma} \sqrt{\lambda_j} t_\gamma] \ni \beta_j\right) \\ &= \mathbb{P}(\kappa_j \ni \beta_j) \end{aligned} \quad (7.26)$$

und damit ist alles gezeigt.  $\square$

### Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Wie gewohnt, betrachten wir als erstes Beispiel die ALM-Form des Szenarios unabhängig und identisch normalverteilter Zufallsvariablen

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^n, \beta := \mu \in \mathbb{R}, \sigma^2 > 0. \quad (7.27)$$

Dann gelten, wie bereits gesehen,

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y}, \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2 \quad \text{und} \quad \lambda_1 = (\mathbf{1}_n^T \mathbf{1}_n)^{-1} = \frac{1}{n}. \quad (7.28)$$

Nach Theorem 7.2 gilt dann, dass

$$\kappa := \left[ \bar{v} - \frac{s}{\sqrt{n}} t_\delta, \bar{v} + \frac{s}{\sqrt{n}} t_\delta \right] \quad (7.29)$$

ein  $\delta$ -Konfidenzintervall für  $\beta$  ist und dieses ist offenbar identisch mit dem bekannten  $\delta$ -Konfidenzintervall für den Erwartungswertparameter der Normalverteilung.

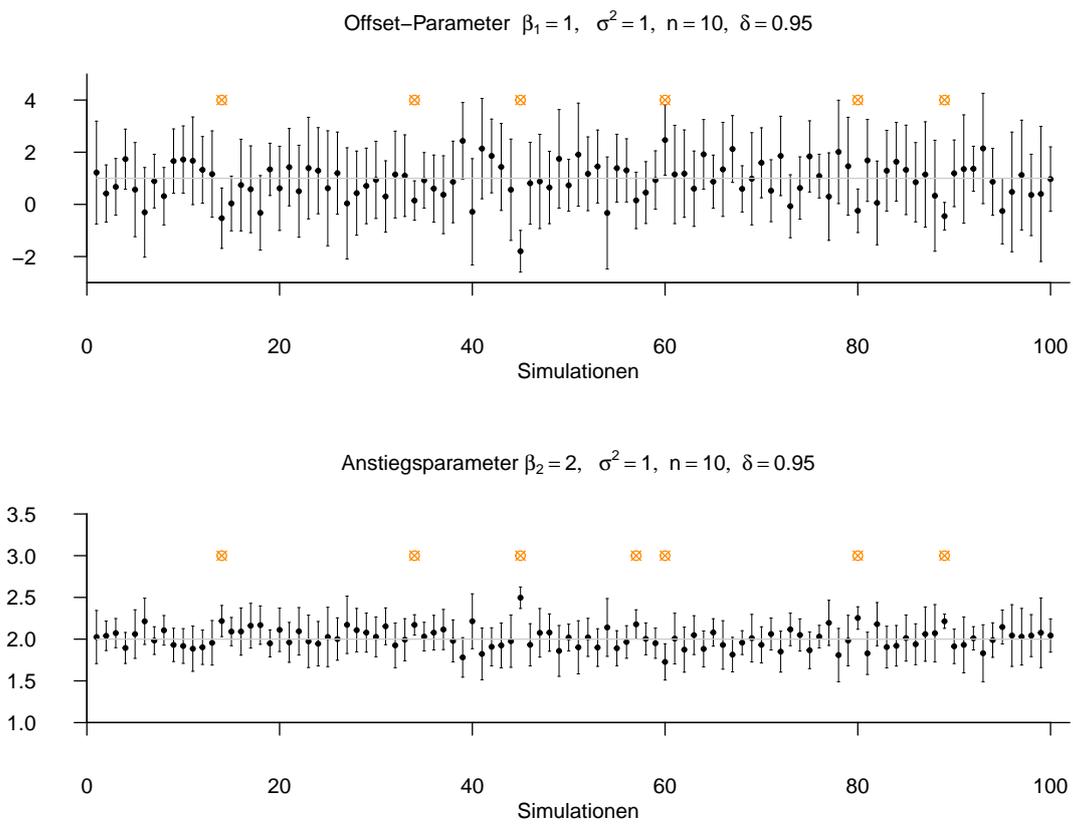


Abbildung 7.3 Konfidenzintervalle für Betaparameterkomponenten bei einfacher linearer Regression.

## Beispiel (2) Einfache lineare Regression

In diesem Beispiel wollen wir nicht auf die spezifische Form der Konfidenzintervalle für den Offset- und Steigungsparameter eingehen, sondern lediglich anhand folgender Simulation an die Frequentistische Bedeutung eines  $\delta$ -Konfidenzintervalls erinnern: Realisierungen von  $\delta$ -Konfidenzintervallen überdecken den wahren, aber unbekanntem Parameterwert mit einer frequentistischen Wahrscheinlichkeit von  $\delta$ .

Abbildung 7.3 zeigt, basierend auf folgendem **R**-Code, dass dies in der konkreten Simulation mit  $\delta = 0.95$  für den Offsetparameter in 94 von 100 und für den Steigungsparameter in 93 von 100 Fällen der Fall ist.

```
# Modellformulierung
library(MASS)
set.seed(0)
ns = 1e2
n = 10
p = 2
x = 1:n
X = matrix(c(rep(1,n),x), ncol = p)
I_n = diag(n)
beta = matrix(c(1,2), nrow = 2)
sigsqr = 1
delta = 0.95
t_delta = qt((1+delta)/2,n-1)
lambda = diag(solve(t(X) %*% X))

# multivariate Normalverteilung
# Zufallszahlengenerator initialisieren
# Anzahl Simulationen
# Anzahl von Datenpunkten
# Anzahl von Betaparametern
# Prädiktorwerte
# n x p Designmatrix
# n x n Einheitsmatrix
# wahre, aber unbekannte Betaparameter
# wahrer, aber unbekannter Varianzparameter
# Konfidenzbedingung
# \Psi^{-1}((1+\delta)/2,n-1)
# \lambda_j Werte

# Simulation
kappa = array(rep(NA, ns*p*p), dim=c(ns,2,2))
beta_hat = matrix(rep(NA,p*ns), nrow = p)
for(i in 1:ns){
  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 = (t(eps_hat) %*% eps_hat)/(n-p)
  for(j in 1:p){
    kappa[i,1,j] = beta_hat[j,i] - sqrt(sigsqr_hat*lambda[j])*t_delta
    kappa[i,2,j] = beta_hat[j,i] + sqrt(sigsqr_hat*lambda[j])*t_delta
  }
}
```

## 7.3 Literaturhinweise

Box (1981) und Zabell (2008) geben einen historischen Überblick zur Entwicklung der T-Statistik und ihrer Verteilung im Kontext der Arbeiten von Student (1908) und Fisher (1925c), Fisher (1925b) und Fisher (1925a). Die Theorie der Konfidenzintervalle geht auf J. Neyman (1935) und Jerzy Neyman (1937) zurück.

## 7.4 Selbstkontrollfragen

1. Geben Sie die Definition einer  $t$ -Zufallsvariable wieder.
2. Geben Sie die Definition einer nichtzentralen  $t$ -Zufallsvariable wieder.
3. Skizzieren Sie die WDFen von  $t$ -Zufallsvariablen mit Freiheitsgradparametern 2, 10 und 30.
4. Skizzieren Sie die WDFen von nichtzentralen  $t$ -Zufallsvariablen mit Nichtzentralitätsparametern 0, 5 und 15.
5. Erläutern Sie den Begriff des Kontrastgewichtsvektors.
6. Geben Sie die Definition der T-Statistik wieder.
7. Erläutern Sie für die T-Statistik die Bedeutung der Wahl von  $c \in \mathbb{R}^p$ .
8. Erläutern Sie für die T-Statistik die Bedeutung der Wahl von  $\beta_0 \in \mathbb{R}^p$ .
9. Wann und warum kann die T-Statistik als Signal-zu-Rauschen-Verhältnis interpretiert werden?
10. Geben Sie das Theorem zur Verteilung der T-Statistik wieder.
11. Geben Sie die Formel für die T-Statistik im ALM-Szenario von unabhängig und identisch normalverteilten Zufallsvariablen wieder.

12. Erläutern Sie den Zusammenhang zwischen der T-Statistik und Cohen's  $d$ .
13. Geben Sie das Theorem zu Konfidenzintervallen für Betaparameterkomponenten wieder.
14. Geben Sie die Formel für das Konfidenzintervall des Erwartungswertparameters im ALM-Szenario von unabhängig und identisch normalverteilten Zufallsvariablen wieder.

## 8 F-Statistiken

In diesem Abschnitt führen wir F-Statistiken vor dem Hintergrund Likelihood-Quotienten-Statistik-basierter Modellvergleiche ein. Die (maximierte oder marginale) Likelihood eines Datensatzes unter einem gegebenen probabilistischen Modell als Modellvergleichskriterium heranzuziehen, ist ein weit verbreitetes Verfahren in der probabilistischen Datenanalyse.

Im Gegensatz zu T-Statistiken kann das Ziel der Berechnung von F-Statistiken damit insbesondere sein, nicht nur Linearkombinationen von Betaparameterschätzwerten probabilistisch zu evaluieren, sondern die Modellanpassung an einen Datensatz insgesamt zu evaluieren. Die Modellvergleichskapazität von F-Statistiken ist allerdings etwas beschränkt, da sich die F-Statistik nur auf ALMs und insbesondere geschachtelte ALMs bezieht, in denen ein Modell Bestandteil eines anderen Modells ist.

F-Statistiken bilden üblicherweise die Grundlage für Hypothesentests im Rahmen varianzanalytischer Verfahren. Der Einsatz von F-Statistiken ist aber *per se* nicht auf Varianzanalysen beschränkt, sondern kann auch bei parametrischen ALM-Designs angebracht sein. Im Folgenden führen wir zunächst den Begriff der Likelihood-Quotienten-Statistik ein und betrachten dann die Definition der F-Statistik vor diesem Hintergrund. Wir schließen mit der Frequentistischen Verteilung der F-Statistik.

### 8.1 Likelihood-Quotienten-Statistiken

Wir definieren den Begriff der Likelihood-Quotienten-Statistik wie folgt.

**Definition 8.1** (Likelihood-Quotienten-Statistik). Gegeben seien zwei parametrische statistische Modelle

$$\mathcal{M}_0 := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_{\theta_0} | \theta_0 \in \Theta_0\}) \text{ und } \mathcal{M}_1 := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_{\theta_1} | \theta_1 \in \Theta_1\}) \quad (8.1)$$

mit identischem Datenraum  $\mathcal{Y}$ , identischer  $\sigma$ -Algebra  $\mathcal{A}$ , aber potentiell distinkten Wahrscheinlichkeitsmaßmengen  $\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta_1}$  und Parameterräumen  $\Theta_1, \Theta_2$ . Es seien weiterhin  $y$  ein Zufallsvektor mit Ergebnisraum  $\mathcal{Y}$  und schließlich  $L_0(y, \theta_0)$  und  $L_1(y, \theta_1)$  die Likelihood-Funktionen von  $\mathcal{M}_0$  bzw.  $\mathcal{M}_1$ , wobei das Argument  $y$  jeweils die Datenabhängigkeit der Likelihood-Funktion anzeigt. Dann wird

$$\Lambda := \frac{\max_{\theta_0 \in \Theta_0} L_0(y, \theta_0)}{\max_{\theta_1 \in \Theta_1} L_1(y, \theta_1)} \quad (8.2)$$

*Likelihood-Quotienten-Statistik* genannt.

•

Eine Likelihood-Quotienten-Statistik setzt die Wahrscheinlichkeitsdichten (oder -massen) eines beobachteten Datensatzes  $y \in \mathcal{Y}$  unter zwei frequentistischen Inferenzmodellen nach Optimierung der jeweiligen Modellparameter ins Verhältnis. Ein hoher Wert der Likelihood-Quotienten-Statistik entspricht einer höheren Wahrscheinlichkeitsdichte des beobachteten Datensatzes  $y \in \mathcal{Y}$  unter  $\mathcal{M}_0$  als unter  $\mathcal{M}_1$  und umgekehrt.

Die Wahrscheinlichkeitsdichten beobachteter Daten nach Modellschätzung unter verschiedenen Modellen zu betrachten, ist ein allgemeines Vorgehen zum Vergleich von Modellen. Letztlich erlaubt dieses Vorgehen, verschiedene wissenschaftliche Theorien über die Entstehung beobachtbarer Daten quantitativ zu vergleichen und die damit verbundene Unsicherheit zu quantifizieren.

Modellvergleiche sind ein zentrales Thema in der Bayesianischen Inferenz, die die Logik von Likelihood-Quotienten-Statistiken zum Beispiel unter den Begriffen der Bayes-Faktoren oder des Bayesianischen Informationskriteriums (BIC) auf allgemeine probabilistische Modelle generalisiert. Allerdings sind, wie hier gesehen, Modellvergleiche auch im Rahmen der frequentistischen Inferenz möglich und sinnvoll, Modellvergleiche sind also kein Alleinstellungsmerkmal der Bayesianischen gegenüber der frequentistischen Inferenz.

Mit dem *reduzierten Modell* und dem *vollständigen Modell* betrachten wir im Folgenden zwei spezielle Formen von  $\mathcal{M}_0$  und  $\mathcal{M}_1$  im Kontext des ALMs.

**Definition 8.2** (Vollständiges und reduziertes Modell.). Für  $p > 1$  mit  $p = p_0 + p_1$  seien

$$X := (X_0 \quad X_1) \in \mathbb{R}^{n \times p} \text{ mit } X_0 \in \mathbb{R}^{n \times p_0} \text{ und } X_1 \in \mathbb{R}^{n \times p_1}, \quad (8.3)$$

sowie

$$\beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^p \text{ mit } \beta_0 \in \mathbb{R}^{p_0} \text{ und } \beta_1 \in \mathbb{R}^{p_1} \quad (8.4)$$

Partitionierungen einer  $n \times p$  Designmatrix und eines  $p$ -dimensionalen Betaparametervektors. Dann sprechen wir von einer *Partitionierung eines Allgemeinen Linearen Modells* und nennen

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (8.5)$$

das *vollständige Modell* und

$$y = X_0\beta_0 + \varepsilon_0 \text{ mit } \varepsilon_0 \sim N(0_n, \sigma_0^2 I_n) \quad (8.6)$$

das *reduzierte Modell*.

•

Man sagt auch, dass das reduzierte Modell im vollständigen Modell *geschachtelt* (engl. *nested*) ist. Die Likelihood-Quotienten-Statistik beim Vergleich eines vollständigen und eines reduzierten Modells hat eine einfache Form, wie folgendes Theorem zeigt.

**Theorem 8.1** (Likelihood-Quotienten-Statistik von vollständigem und reduziertem Modell). Für  $p = p_0 + p_1$  mit  $p > 1$  sei eine Partitionierung eines vollständigen ALMs gegeben und es seien  $\hat{\sigma}^2$  und  $\hat{\sigma}_0^2$  die Maximum-Likelihood-Schätzer des Varianzparameters unter vollständigem bzw. reduziertem Modell. Weiterhin seien die zwei parametrischen

statistischen Modelle  $\mathcal{M}_0$  und  $\mathcal{M}_1$  in der Definition der Likelihood-Quotienten-Statistik durch das reduzierte Modell und das vollständige Modell gegeben. Dann gilt

$$\Lambda = \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{\frac{n}{2}} \quad (8.7)$$

◦

*Beweis.* Wir erinnern zunächst daran, dass die Maximum-Likelihood-Schätzer des Varianzparameters durch

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - X\hat{\beta})^\top (\mathbf{y} - X\hat{\beta}) \quad \text{und} \quad \hat{\sigma}_0^2 = \frac{1}{n} (\mathbf{y} - X_0\hat{\beta}_0)^\top (\mathbf{y} - X_0\hat{\beta}_0) \quad (8.8)$$

gegeben sind, wobei  $\hat{\beta}$  und  $\hat{\beta}_0$  die Maximum-Likelihood-Schätzer der Betaparameter unter vollständigem bzw. reduziertem Modell bezeichnen. Weiterhin halten wir fest, dass die Likelihood-Funktion des vollständigen Modells an der Stelle der Maximum-Likelihood-Schätzer den folgenden Wert annimmt

$$\begin{aligned} L_1(\mathbf{y}, \hat{\beta}, \hat{\sigma}^2) &= \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n}} \cdot \exp\left(-\frac{1}{2\hat{\sigma}^2} (\mathbf{y} - X\hat{\beta})^\top (\mathbf{y} - X\hat{\beta})\right) \\ &= \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n}} \cdot \exp\left(-\frac{n}{2} \frac{(\mathbf{y} - X\hat{\beta})^\top (\mathbf{y} - X\hat{\beta})}{(\mathbf{y} - X\hat{\beta})^\top (\mathbf{y} - X\hat{\beta})}\right) \\ &= \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n}} \cdot e^{-\frac{n}{2}}, \end{aligned} \quad (8.9)$$

und analog, für die Likelihood-Funktion des reduzierten Modells an der Stelle der Maximum-Likelihood-Schätzer

$$L_0(\mathbf{y}, \hat{\beta}_0, \hat{\sigma}_0^2) = \frac{1}{\sqrt{(2\pi)^n (\hat{\sigma}_0^2)^n}} \cdot e^{-\frac{n}{2}}. \quad (8.10)$$

Damit ergibt sich dann aber

$$\Lambda = \frac{\max_{\theta_0 \in \Theta_0} L_0(\mathbf{y}, \theta_0)}{\max_{\theta_1 \in \Theta_1} L_1(\mathbf{y}, \theta_1)} = \frac{L_0(\mathbf{y}, \hat{\beta}_0, \hat{\sigma}_0^2)}{L_1(\mathbf{y}, \hat{\beta}, \hat{\sigma}^2)} = \frac{\sqrt{(2\pi)^n (\hat{\sigma}^2)^n} \cdot e^{-\frac{n}{2}}}{\sqrt{(2\pi)^n (\hat{\sigma}_0^2)^n} \cdot e^{-\frac{n}{2}}} = \sqrt{\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^n} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}}. \quad (8.11)$$

□

## 8.2 Definition und Verteilung

Wir definieren nun die F-Statistik vor dem Hintergrund eines vollständigen und eines reduzierten Modells.

**Definition 8.3** (F-Statistik). Gegeben seien das ALM

$$\mathbf{y} = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (8.12)$$

wobei  $X \in \mathbb{R}^{n \times p}$ , und mit  $p = p_0 + p_1$  eine Partitionierung

$$\begin{aligned} X &= (X_0 \quad X_1), \quad X_0 \in \mathbb{R}^{n \times p_0}, \quad X_1 \in \mathbb{R}^{n \times p_1} \quad \text{und} \\ \beta &= \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \beta_0 \in \mathbb{R}^{p_0}, \quad \beta_1 \in \mathbb{R}^{p_1}. \end{aligned} \quad (8.13)$$

Weiterhin seien  $\hat{\beta}$  und  $\hat{\beta}_0$  die Betaparameterschätzer des vollständigen bzw. reduzierten Modells sowie  $\hat{\varepsilon}$  und  $\hat{\varepsilon}_0$  die Residuenvektoren des vollständigen bzw. reduzierten Modells:

$$\begin{aligned}\hat{\varepsilon} &:= y - X\hat{\beta} & \text{mit} & \quad \hat{\beta} := (X^T X)^{-1} X^T y \\ \hat{\varepsilon}_0 &:= y - X_0\hat{\beta}_0 & \text{mit} & \quad \hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T y.\end{aligned}\tag{8.14}$$

Dann ist die F-Statistik definiert als

$$F := \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)}.\tag{8.15}$$

•

Der Zähler der F-Statistik

$$\frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_1}\tag{8.16}$$

misst, inwieweit die  $p_1$  Regressoren in  $X_1$  die Residualquadratsumme reduzieren und zwar im Verhältnis zur Anzahl dieser Regressoren. Das heißt, dass bei gleicher Residualquadratsummenreduktion (und gleichem Nenner) ein größerer  $F$ -Wert resultiert, wenn diese durch weniger zusätzliche Regressoren erreicht wird, also  $p_1$  klein ist (und umgekehrt). Im Sinne der Anzahl der Spalten von  $X$  und der entsprechenden Komponenten von  $\beta$  favorisiert die  $F$ -Statistik also weniger “komplexe” Modelle.

Für den Nenner der F-Statistik gilt

$$\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p} = \hat{\sigma}^2,\tag{8.17}$$

wobei  $\hat{\sigma}^2$  hier der aufgrund des vollständigen Modells geschätzte Schätzer von  $\sigma^2$  ist. Wenn die Daten tatsächlich unter dem reduzierten Modell generiert werden, so kann das vollständige Modell dies durch  $\hat{\beta}_1 \approx 0_{p_1}$  abbilden und erreicht eine ähnliche Varianzschätzung wie das reduzierte Modell. Wenn die Daten tatsächlich unter dem vollständigem Modell generiert werden, so ist  $\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)$  ein besserer Schätzer von  $\sigma^2$  als  $\hat{\varepsilon}_0^T \hat{\varepsilon}_0/(n-p)$ , da sich für ersteren Datenvariabilität, die nicht durch die  $p_0$  Regressoren in  $X_0$  erklärt wird, in der Schätzung von  $\sigma^2$  widerspiegeln würde. Der Nenner der F-Statistik ist also in beiden Fällen der sinnvollere Schätzer von  $\sigma^2$ .

Zusammengenommen misst die F-Statistik also die Residualquadratsummenreduktion durch die  $p_1$  Regressoren in  $X_1$  gegenüber den  $p_0$  Regressoren in  $X_0$  pro Datenvariabilitäts ( $\sigma^2$ )- und Regressor ( $p_1$ )-Einheit.

## Beispiel (1) Einfache lineare Regression

Exemplarisch evaluiert untenstehender **R**-Code die F-Statistik im Kontext folgender Partitionierung des Modells der einfachen linearen Regression

$$X = (X_0 \quad X_1), \quad X_0 := \mathbf{1}_n, \quad X_1 := (x_1, \dots, x_n)^T, \quad \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}\tag{8.18}$$

und zwar einmal für den Fall, dass  $\beta = (1, 0)^T$ , also das reduzierte Modell das wahre, aber unbekannte datenerzeugende Modell ist, und einmal für den Fall, dass  $\beta = (1, 0.5)^T$ ,

also das vollständige Modell das wahre, aber unbekannte datenerzeugende Modell ist. Im Sinne obiger Diskussion ergibt sich im ersten Fall eine F-Statistik nahe Null, im zweiten Fall dagegen eine hohe F-Statistik.

```
# Modellformulierung
library(MASS)
nmod = 2
n = 10
p = 2
p_0 = 1
p_1 = 1
p = p_0 + p_1
x = 1:n
X = matrix(c(rep(1,n),x), nrow = n)
X_0 = X[,1]
I_n = diag(n)
beta = matrix(c(1,0,
                1,0.5), nrow = 2)
nscn = ncol(beta)
sigsqr = 1

# multivariate Normalverteilung
# Anzahl Modelle
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Anzahl Betaparameter reduziertes Modell
# Anzahl zusätzlicher Betaparameter vollständiges Modell
# Anzahl Betaparameter im vollständigem Modell
# Prädiktorwerte
# Designmatrix des vollständigen Modells
# Designmatrix des reduzierten Modells
# Einheitsmatrix

# wahre, aber unbekannte Betaparameter
# Anzahl wahrer, aber unbekannter Hypothesenszenarien
# wahrer, aber unbekannter Varianzparameter

# Modellsimulation und Evaluierung
Eff = matrix(rep(NA, nscn), nrow = nscn)
for(s in 1:nscn){
  y = mvrnorm(1, X %>% beta[,s], sigsqr*I_n)
  beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y
  beta_hat = solve(t(X) %*%X )%*%t(X) %*%y
  eps_0_hat = y-X_0%*%beta_hat_0
  eps_hat = y-X%*%beta_hat
  eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat
  eps_eps_hat = t(eps_hat) %*% eps_hat
  Eff[s] = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/
            (eps_eps_hat/(n-p)))}

# F-Statistik Realisierungsarray
# Szenarieniterationen
# Datenrealisierung
# Betaparameterschätzer reduziertes Modell
# Betaparameterschätzer vollständiges Modell
# Residuenvektor reduziertes Modell
# Residuenvektor vollständiges Modell
# RQS reduziertes Modell
# RQS vollständiges Modell
# F-Statistik
```

F-Statistik für beta\_1 = 0\_{p\_1}: 0.034497  
 F-Statistik für beta\_1 != 0\_{p\_1}: 20.72964

Die Likelihood-Quotienten-Statistik und die F-Statistik von vollständigem und reduziertem Modell sind ineinander überführbar. Dies ist die zentrale Aussage folgenden Theorems.

**Theorem 8.2** (F-Statistik und Likelihood-Quotienten-Statistik.). *Es sei die Partitionierung eines ALMs in ein vollständiges und ein reduziertes Modell gegeben und  $F$  und  $\Lambda$  seien die entsprechenden F- und Likelihood-Quotienten-Statistiken. Dann gilt*

$$F = \frac{n-p}{p_1} \left( \Lambda^{-\frac{2}{n}} - 1 \right). \quad (8.19)$$

o

*Beweis.* Wir erinnern zunächst daran, dass die Maximum-Likelihood-Schätzer des Varianzparameters durch

$$\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})^T (y - X\hat{\beta}) = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n} \quad \text{und} \quad \hat{\sigma}_0^2 = \frac{1}{n} (y - X_0\hat{\beta}_0)^T (y - X_0\hat{\beta}_0) = \frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0}{n} \quad (8.20)$$

gegeben sind. Mit der Definition der F-Statistik und der Form der Likelihood-Quotienten-Statistik für den Vergleich von reduziertem und vollständigem Modell

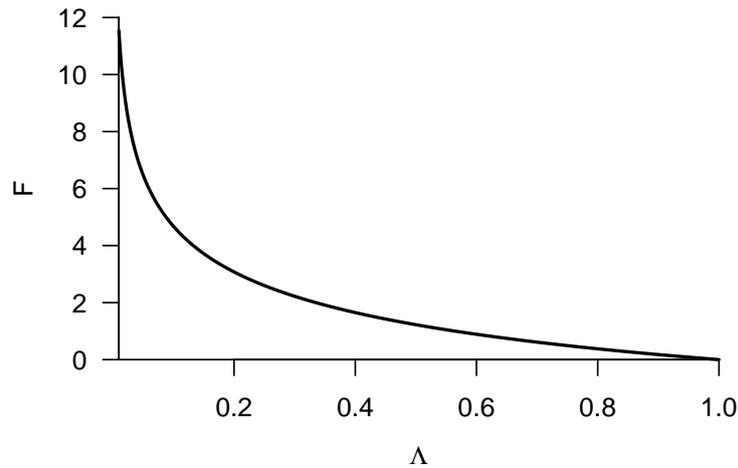
$$\Lambda = \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{\frac{n}{2}} \quad (8.21)$$

ergibt sich dann

$$\begin{aligned}
 F &= \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \\
 &= \frac{n(\hat{\sigma}_0^2 - \hat{\sigma}^2)/p_1}{n\hat{\sigma}^2/(n-p)} \\
 &= \frac{n-p}{p_1} \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \\
 &= \frac{n-p}{p_1} \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1 \right) \\
 &= \frac{n-p}{p_1} \left( \Lambda^{-\frac{2}{n}} - 1 \right) .
 \end{aligned} \tag{8.22}$$

□

Zwischen der F-Statistik und der Likelihood-Quotienten-Statistik besteht also ein nichtlinearer, reziproker Zusammenhang, den wir für  $n = 12$ ,  $p = 2$  und  $p_1 = 1$  in Abbildung 8.1 visualisieren. Man beachte, dass für  $\Lambda = 1$  gilt, dass  $F = 0$ . Ein Wert von  $F = 0$  impliziert also, dass das reduzierte Modell gegenüber dem vollständigen Modell im Lichte eines beobachteten Datensatzes die gleiche Plausibilität besitzt.



**Abbildung 8.1** Zusammenhang von F- und Likelihood-Quotienten-Statistik.

Wir dokumentieren die Frequentistische Verteilung der F-Statistik in folgendem Theorem, dessen Beweis wir allerdings nur andeuten werden.

**Theorem 8.3** (Verteilung der F-Statistik). *Gegeben seien das ALM*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n), \tag{8.23}$$

wobei  $X \in \mathbb{R}^{n \times p}$ , und mit  $p = p_0 + p_1$  eine Partitionierung

$$\begin{aligned}
 X &= (X_0 \ X_1), \ X_0 \in \mathbb{R}^{n \times p_0}, \ X_1 \in \mathbb{R}^{n \times p_1} \text{ und} \\
 \beta &= \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \ \beta_0 \in \mathbb{R}^{p_0}, \ \beta_1 \in \mathbb{R}^{p_1}.
 \end{aligned} \tag{8.24}$$

Schließlich sei ein Vektor  $c$  gegeben durch

$$c := \begin{pmatrix} 0_{p_0} \\ 1_{p_1} \end{pmatrix} \in \mathbb{R}^p. \quad (8.25)$$

Dann gilt

$$F \sim f(\delta, p_1, n - p) \text{ mit } \delta := \frac{c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta}{\sigma^2} \quad (8.26)$$

◦

*Beweis.* Wir haben bereits gesehen, dass folgende Funktion des Varianzparameterschätzers einer Chi-Quadrat-Verteilung folgt (siehe Einheit (6) in *Allgemeines Lineares Modell*):

$$U_2 = \frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n - p). \quad (8.27)$$

Des Weiteren lässt sich zeigen, dass folgende Funktion der Varianzparameterschätzer von vollständigem und reduziertem Modell einer nichtzentralen Chi-Quadrat-Verteilung folgt:

$$U_1 = \frac{n - p_0}{\sigma^2} \hat{\sigma}_0^2 - \frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2 \left( \frac{c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta}{\sigma^2}, p - p_0 \right). \quad (8.28)$$

Definitionsgemäß ist der Quotient aus der Zufallsvariable  $U_1$ , geteilt durch ihren Freiheitsgradparameter, und der Zufallsvariable  $U_2$ , geteilt durch ihren Freiheitsgradparameter, eine nichtzentrale  $f$ -Zufallsvariable

$$F = \frac{U_1 / (p - p_0)}{U_2 / (n - p)} \sim f(\delta, p - p_0, n - p), \quad (8.29)$$

wobei der Nichtzentralitätsparameter der nichtzentralen  $f$ -Verteilung durch den Nichtzentralitätsparameter der nichtzentralen Chi-Quadrat-Verteilung gegeben ist und sich die  $F$ -Statistik wie folgt ergibt:

$$\begin{aligned} F &= \frac{\left( \frac{n - p_0}{\sigma^2} \hat{\sigma}_0^2 - \frac{n - p}{\sigma^2} \hat{\sigma}^2 \right) / (p - p_0)}{\frac{n - p}{\sigma^2} \hat{\sigma}^2 / (n - p)} = \frac{\frac{1}{\sigma^2} \left( (n - p_0) \hat{\sigma}_0^2 - (n - p) \hat{\sigma}^2 \right) / (p - p_0)}{\frac{1}{\sigma^2} (n - p) \hat{\sigma}^2 / (n - p)} \\ &= \frac{\left( (n - p_0) \hat{\sigma}_0^2 - (n - p) \hat{\sigma}^2 \right) / p_1}{(n - p) \hat{\sigma}^2 / (n - p)} = \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}) / p_1}{\hat{\varepsilon}^T \hat{\varepsilon} / (n - p)}. \end{aligned} \quad (8.30)$$

□

Man beachte, dass die  $F$ -Statistik eine Funktion der Parameterschätzer,  $\delta$  dagegen eine Funktion der wahren, aber unbekannt Parameter ist. Wie die Verteilung der  $T$ -Statistik kann auch die Verteilung der  $F$ -Statistik für die Evaluation von frequentistischen Konfidenzintervallen und Hypothesentests genutzt werden. Insbesondere letzteren Aspekt verdeutlichen wir in Kapitel 10 und Kapitel 11.

## Beispiel (1) Einfache lineare Regression

Exemplarisch evaluieren wir mithilfe untenstehenden **R**-Codes die Verteilung der  $F$ -Statistik im Kontext der Partitionierung des Modells der einfachen linearen Regression und zwar erneut einmal für den Fall, dass  $\beta = (1, 0)^T$ , also das reduzierte Modell das wahre, aber unbekannt datenerzeugende Modell ist, und einmal für den Fall, dass  $\beta = (1, 0.5)^T$ , also das vollständige Modell das wahre, aber unbekannt datenerzeugende Modell ist. Abbildung 8.2 A und B visualisieren die resultierenden Verteilungen.

```

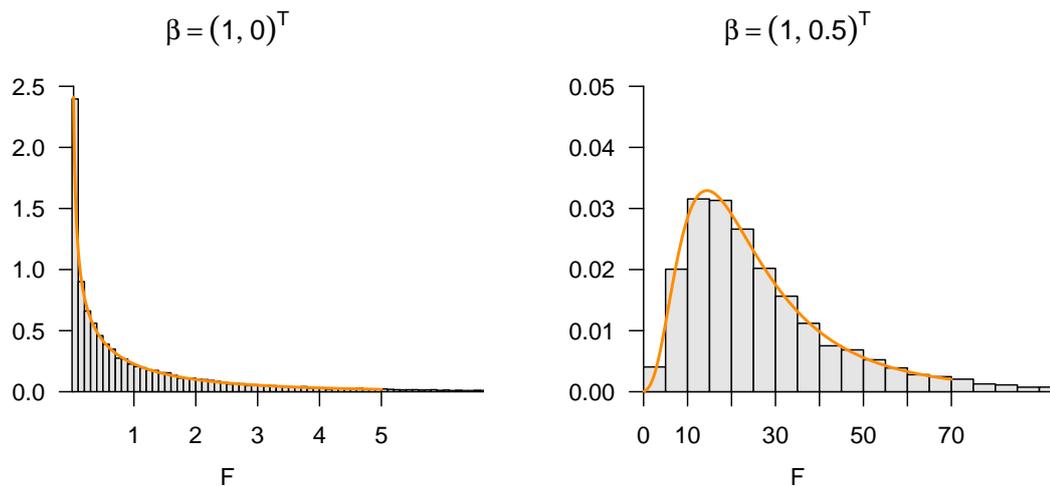
# Modellformulierung
library(MASS)
nmod = 2
n = 10
p_0 = 1
p_1 = 1
p = p_0 + p_1
x = 1:n
X = matrix(c(rep(1,n),x), nrow = n)
X_0 = X[,1]
I_n = diag(n)
beta = matrix(c(1,0,
                1,0.5), nrow = 2)
nscn = ncol(beta)
sigsqr = 1
c = matrix(c(0,1), nrow = 2)

# multivariate Normalverteilung
# Anzahl Modelle
# Anzahl Datenpunkte
# 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
# Einheitsmatrix

# wahre, aber unbekannte Betaparameter
# Anzahl wahrer, aber unbekannter Hypothesenszenarien
# wahrer, aber unbekannter Varianzparameter
# Vektor

# Frequentistische Simulation
nsim = 1e4
delta = rep(NA,nscn)
Eff = matrix(rep(NA,n), nrow = nscn)
for(s in 1:nscn){
  delta[s] = (t(t(c)%*%beta[,s])%*%
             solve(t(c)%*%solve(t(X)%*%X)%*%c) %*%
             (t(c)%*%beta[,s])/sigsqr)
  for(i in 1:nsim){
    y = mvrnorm(1, X %*%beta[,s], sigsqr*I_n)
    beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y
    beta_hat = solve(t(X) %*%X )%*%t(X) %*%y
    eps_0_hat = y-X_0%*%beta_hat_0
    eps_hat = y-X%*%beta_hat
    eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat
    eps_eps_hat = t(eps_hat) %*% eps_hat
    Eff[s,i] = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/
               (eps_eps_hat/(n-p)))
  }
}

```



**Abbildung 8.2** Exemplarische F-Statistik Verteilungen unter reduziertem und vollständigem wahren, aber unbekanntem datenerzeugendem Modell bei einfacher linearer Regression.

### 8.3 Literaturhinweise

Die Popularität von F-Statistiken, insbesondere im Kontext der Varianzanalyse, wird allgemein auf Fisher (1925a) zurückgeführt. Seal (1967) gibt einen historischen Überblick. Likelihood-Quotienten-Statistiken werden insbesondere von J. Neyman und Pearson (1928) und Wilks (1938) betrachtet. Lehmann (2011) gibt einen integrierten historischen Überblick zu beiden Ansätzen. Die hier diskutierte Äquivalenz von Likelihood-Quotienten-Statistik und F-Statistik basiert auf der Darstellung in Seber und Lee (2003).

## 8.4 Selbstkontrollfragen

1. Geben Sie die Definition einer  $f$ -Zufallsvariable wieder.
2. Geben Sie die Definition einer nichtzentralen  $f$ -Zufallsvariable wieder.
3. Skizzieren Sie die WDF einer  $f$ -Verteilung mit  $n_1 = 4, n_2 = 26$ .
4. Skizzieren Sie die WDF einer nichtzentralen  $f$ -Verteilung mit  $\delta = 4, n_1 = 2, n_2 = 26$ .
5. Geben Sie die Definition der Likelihood-Quotienten-Statistik wieder.
6. Erläutern Sie die Definition der Likelihood-Quotienten-Statistik.
7. Geben Sie die Definition von vollständigem und reduziertem ALM wieder.
8. Geben Sie das Theorem zum Likelihood-Quotienten von vollständigem und reduzierten ALM wieder.
9. Definieren Sie die F-Statistik.
10. Erläutern Sie den Zähler der F-Statistik.
11. Erläutern Sie den Nenner der F-Statistik.
12. Erläutern Sie die F-Statistik.
13. Geben Sie das Theorem zum Zusammenhang von F-Statistik und Likelihood-Quotienten-Statistik wieder.
14. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

# 9 T-Tests

## 9.1 Einstichproben-T-Tests

### Anwendungsszenario

Das Anwendungsszenario eines Einstichproben-T-Tests ist bekanntlich dadurch gekennzeichnet, dass  $n$  univariate Datenpunkte einer Stichprobe (Gruppe) randomisierter experimenteller Einheiten betrachtet werden, von denen angenommen wird, dass sie Realisierungen von  $n$  unabhängigen und identisch normalverteilten Zufallsvariablen sind. Hinsichtlich der identischen univariaten Normalverteilungen  $N(\mu, \sigma^2)$  dieser Zufallsvariablen wird angenommen, dass sowohl der Erwartungswertparameter  $\mu$  als auch der Varianzparameter  $\sigma^2$  unbekannt sind. Schließlich wird vorausgesetzt, dass ein Interesse an einem inferentiellen Vergleich des unbekanntem Erwartungswertparameters  $\mu$  mit einem vorgegebenen Wert  $\mu_0$  (z.B.  $\mu_0 := 0$ ) besteht.

### Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel betrachten wir die Analyse von BDI-Differenzwerten in einer Pre-Post-Interventionsstudie für eine Gruppe von  $n = 12$  Patient:innen, wie in Tabelle 9.1 dargestellt. Die ersten beiden Spalten dieser Tabelle listen die patientenspezifische BDI-Werte vor (*PreBDI*) und nach (*PostBDI*) der psychotherapeutischen Intervention, die dritte Spalte *dBDI* zeigt die entsprechenden Pre-Post-Differenzwerte. Ein positiver Wert entspricht hier einer Verbesserung der Depressionssymptomatik und ein negativer Wert einer Verschlechterung der Depressionssymptomatik.

**Tabelle 9.1** Pre- und Post-Interventions-BDI-Werte

PreBDI	PosBDI	dBDI
29	29	0
32	27	5
28	27	1
36	22	14
32	25	7
28	28	0
33	31	2
33	27	6
33	28	5
30	27	3
36	23	13
32	26	6

Bei der Anwendung eines Einstichproben-T-Tests auf die *dBDI*-Werte dieses Datensatzes nehmen wir also an, dass die *dBDI*-Daten Realisierungen von  $n = 12$  unabhängig normalverteilten Zufallsvariablen  $y_i \sim N(\mu, \sigma^2)$  sind. Wir nehmen weiterhin an, dass wir

daran interessiert sind, unsere Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekanntem, Erwartungswertparameters  $\mu$  mit einem Vergleichswert  $\mu_0$  im Sinne eines Hypothesentests zu quantifizieren.

Unabhängig von diesem inferenzstatistischen Vorgehen betrachten wir zunächst die deskriptiven Statistiken der dBDI-Daten, wie in Tabelle 9.2 dargestellt. Es fällt insbesondere auf, dass das Stichprobenmittel im Vergleich zur Standardabweichung relativ klein ist. Im Gruppenmittel unterscheiden sich PreBDI und PostBDI also zwar in positiver Richtung, was eine Verringerung der Depressionssymptomatik andeutet, allerdings streuen die Daten über Patient:innen auch deutlich, wie auch bereits aus Tabelle 9.1 ersichtlich.

**Tabelle 9.2** Deskriptivstatistiken der Pre-Post BDI Differenzwerte

	n	Max	Min	Median	Mean	Var	Std
F2F	12	14	0	5	5.17	20.88	4.57

## Modellformulierung

Wir definieren nun das Einstichproben-T-Test-Modell wie folgt.

**Definition 9.1** (Einstichproben-T-Test-Modell).  $y_i, i = 1, \dots, n$  seien Zufallsvariablen, die die  $n$  Datenpunkte eines Anwendungsszenarios für den Einstichproben-T-Test modellieren. Dann hat das *Einstichproben-T-Test-Modell* die strukturelle Form

$$y_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.1)$$

die Datenverteilungsform

$$y_i \sim N(\mu, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.2)$$

und für den Datenvektor  $y = (y_1, \dots, y_n)^T$  die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ und } \sigma^2 > 0. \quad (9.3)$$

•

Das Modell des Einstichproben-T-Tests ist offenbar mit dem dem Modell unabhängiger und identisch normalverteilter Zufallsvariablen identisch (vgl. Kapitel 5). Die Äquivalenz von struktureller, Datenverteilungs- und Designmatrixform des Einstichproben-T-Test-Modells wurde in Kapitel 5 bereits ausführlich diskutiert. Die Simulation von Daten basierend auf dem Einstichproben-T-Test-Modell hat dementsprechend die gleiche Form wie die Simulation unabhängig und identisch normalverteilter Zufallsvariablen. Der untenstehende **R**-Code demonstriert dies.

```
# Modellformulierung
library(MASS)                # multivariate Normalverteilung
n = 12                       # Anzahl von Datenpunkten
p = 1                        # Anzahl von Betaparameter
X = matrix(rep(1,n), nrow = n) # n x p 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, sigsq*I_n) # eine Realisierung des n-dimensionalen ZVs y

```

## Modellschätzung

Da die Form des Einstichproben-T-Test-Modells mit dem Szenario unabhängig und identisch normalverteilter Zufallsvariablen identisch ist, trifft dies auch auf die entsprechenden Beta- und Varianzparameterschätzer zu. Es ergibt sich also folgendes Theorem, das bereits in Kapitel 6 bewiesen wurde.

**Theorem 9.1** (Parameterschätzer im Einstichproben-T-Test-Modell). *Gegeben sei die Designmatrixform des Einstichproben-T-Test-Modells. Dann ergeben sich für den Betaparameterschätzer*

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y} \quad (9.4)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2. \quad (9.5)$$

◦

## Modellevaluation

Basierend auf Definition 7.1 formulieren wir nun die T-Teststatistik für das Einstichproben-T-Test Szenario und geben ihre frequentistische Verteilung an.

**Theorem 9.2** (T-Teststatistik des Einstichproben-T-Tests). *Gegeben sei die Designmatrixform des Einstichproben-T-Test-Modells. Dann ergibt sich für die T-Teststatistik mit*

$$c := 1 \text{ und } c^T \beta_0 =: \mu_0, \quad (9.6)$$

dass

$$T = \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right) \quad (9.7)$$

und es gilt

$$T \sim t(\delta, n-1) \text{ mit } \delta = \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right). \quad (9.8)$$

◦

*Beweis.* Mit dem Theorem zur Verteilung der T-Statistik (siehe Theorem 7.1) gilt

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y} - 1^T \mu_0}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right). \quad (9.9)$$

Weiterhin gilt mit demselben Theorem

$$\delta = \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} = \frac{1^T \mu - 1^T \mu_0}{\sqrt{\sigma^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right). \quad (9.10)$$

□

Die Formen der T-Teststatistik und ihre Verteilung im ALM-Spezialfall des Einstichproben-T-Tests sind also natürlicherweise mit den entsprechenden Formen im ALM-freien Kontext identisch. Die entsprechende Theorie zu Konfidenzintervallen und der Kontrolle des Testumfangs bei Einstichproben-T-Tests sowie der Gebrauch der Testgütefunktion zur Evaluation der Testtrennschärfe (Power) folgt also analog.

## Anwendungsbeispiel

Folgender **R**-Code demonstriert die Evaluation eines 95%-Konfidenzintervalls für den Erwartungswertparameter  $\mu$  sowie Durchführung eines zweiseitigen Einstichproben-T-Tests mit einfacher Nullhypothese  $\Theta_0 := \{0\}$  und Signifikanzlevel  $\alpha_0 := 0.05$  für das oben skizzierte Anwendungsbeispiel.

```
# Datenanalyse
D = read.csv("_data/09_T-Tests.csv") # Laden des Datensatzes
y = D$BDDI[D$COND == "F2F"] # Pre-Post-Differenzwerte
n = length(y) # Anzahl Datenpunkte
p = 1 # Anzahl Betaparameter
c = 1 # Kontrastgewichtsvektor
mu_0 = 0 # Nullhypotheseparameter
delta = 0.95 # Konfidenzlevel
alpha_0 = 0.05 # Signifikanzlevel
X = matrix(rep(1,n), nrow = n) # Einstichproben-T-Test 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
t_delta = qt((1+delta)/2, n-1) # \Psi^{-1}((1+\delta)/2, n-1)
lambda = diag(solve(t(X) %*% X)) # \lambda_j Werte
kappa_u = beta_hat - sqrt(sigsqr_hat*lambda)*t_delta # untere Konfidenzintervallgrenze
kappa_o = beta_hat + sqrt(sigsqr_hat*lambda)*t_delta # obere Konfidenzintervallgrenze
t_num = t(c) %*% beta_hat - mu_0 # Zähler der Einstichproben-T-Teststatistik
t_den = sqrt(sigsqr_hat %*% t(c) %*% solve(t(X) %*% X) %*% c) # Nenner der Einstichproben-T-Teststatistik
t = t_num/t_den # Wert der Einstichproben-T-Teststatistik
pval = 2*(1 - pt(abs(t), n-1)) # p-Wert bei zweiseitigem Einstichproben-T-Test
k_alpha_0 = qt(1-alpha_0/2, n-1) # kritischer Wert
if(abs(t) > k_alpha_0){phi = 1} else {phi = 0} # Einstichproben-T-Test
```

```
Betaparameterschätzer : 5.17
95%-Konfidenzintervall : 2.26 8.07
Varianzparameterschätzer : 20.88
Signifikanzlevel : 0.05
Kritischer Wert : 2.2
Einstichproben-T-Teststatistik : 3.92
Testergebnis : 1
p-Wert : 0
```

Die Nullhypothese wird in diesem Fall also abgelehnt, da der Wert der T-Statistik von  $T = 3.92$  den kritischen Wert von  $k_{0.05} = 2.20$  übersteigt. Das 95%-Konfidenzintervall für den wahren, aber unbekanntem Erwartungswertparameter ist  $[2.26, 8.07]$ , überdeckt also den Nullhypotheseparameterwert  $\mu_0 = 0$  nicht.

## 9.2 Zweistichproben-T-Tests

### Anwendungsszenario

Das Anwendungsszenario eines Zweistichproben-T-Tests für unabhängige Stichproben ist bekanntlich dadurch gekennzeichnet, dass insgesamt  $n$  univariate Datenpunkte zweier Stichproben (Gruppen) randomisierter experimenteller Einheiten betrachtet werden. Es wird dabei insbesondere angenommen, dass die  $n_1$  univariaten Datenpunkte der ersten Gruppe Realisierungen von  $n_1$  unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter  $\mu_1$  und Varianzparameter  $\sigma^2$  sind und dass die  $n_2$  univariaten Datenpunkte der zweiten Gruppe Realisierungen von  $n_2$  unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter  $\mu_2$  und Varianzparameter  $\sigma^2$  sind.

Es wird also davon ausgegangen, dass sich die wahren, aber unbekanntes Erwartungswertparameter beider Gruppen von Zufallsvariablen unterscheiden können, die Varianzparameter beider Gruppen dagegen werden identisch sind. Schließlich wird vorausgesetzt, dass ein Interesse am inferentiellen Vergleich der unbekanntes Erwartungswertparameter  $\mu_1$  und  $\mu_2$  besteht, so zum Beispiel ihrer Gleichheit ( $\mu_1 = \mu_2$ ) oder Verschiedenheit ( $\mu_1 \neq \mu_2$ ).

### Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel betrachten wir die Analyse von BDI-Differenzwerten in einer Pre-Post-Interventionsstudie für zwei Gruppen von je 12 Patient:innen in unterschiedlichen Therapiesettings, wie in Tabelle 9.3 dargestellt. Die erste Spalte der Tabelle (**Setting**) listet das patientenspezifische Therapiesetting (F2F: face-to-face, ONL: online) auf. Die zweite Spalte der Tabelle (**dBDI**) listet die entsprechenden patientenspezifischen BDI-Differenzwerte auf. Positive Werte entsprechen hier erneut einer Abnahme der Depressionssymptomatik, negative Werte einer Zunahme der Depressionssymptomatik.

**Tabelle 9.3** Pre-Post-BDI Differenzwerte für zwei Stichproben

COND	dBDI
F2F	0
F2F	5
F2F	1
F2F	14
F2F	7
F2F	0
F2F	2
F2F	6
F2F	5
F2F	3
F2F	13
F2F	6
ONL	5
ONL	-1
ONL	7
ONL	4
ONL	7
ONL	10
ONL	7

**Tabelle 9.3** Pre-Post-BDI Differenzwerte für zwei Stichproben

COND	dBDI
ONL	6
ONL	11
ONL	10
ONL	5
ONL	-2

Zu Anwendung eines Zweistichproben-T-Tests auf die dBDI Daten nehmen wir an, dass die 12 Datenpunkte der F2F-Therapiegruppe Realisierungen von  $n_1 = 12$  unabhängig und identisch normalverteilten Zufallsvariablen  $y_{1j} \sim N(\mu_1, \sigma^2)$  mit  $j = 1, \dots, n_1$  sind und dass die 12 Datenpunkte der ONL-Therapiegruppe Realisierungen von  $n_2 = 12$  unabhängig und identisch normalverteilten Zufallsvariablen  $y_{2j} \sim N(\mu_2, \sigma^2)$  mit  $j = 1, \dots, n_2$  sind.

Unabhängig von dem unten beschriebenen inferenzstatistischen Vorgehen betrachten wir auch hier zunächst die deskriptiven Statistiken der Therapiesetting-spezifischen dBDI Werte. Diese sind in Tabelle 9.4 aufgeführt.

**Tabelle 9.4** Deskriptivstatistiken der Pre-Post BDI Differenzwerte bei unterschiedlichen Therapiesettings

	n	Max	Min	Median	Mean	Var	Std
F2F	12	14	0	5.0	5.17	20.88	4.57
ONL	12	11	-2	6.5	5.75	16.20	4.03

## Modellformulierung

Mit dem Index  $i$  für die Gruppen und dem Index  $j$  für die experimentellen Einheiten in jeder Gruppe definieren wir das Zweistichproben-T-Test-Modell wie folgt.

**Definition 9.2** (Zweistichproben-T-Test-Modell).  $y_{ij}$  mit  $i = 1, 2$  und  $j = 1, \dots, n_i$  seien Zufallsvariablen, die die Datenpunkte eines Anwendungsszenarios für den Zweistichproben-T-Test modellieren. Dann hat das *Zweistichproben-T-Test-Modell* die strukturelle Form

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.11)$$

die Datenverteilungsform

$$y_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, \dots, n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.12)$$

und für den Datenvektor  $y = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2})^T$  und  $n := n_1 + n_2$  die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (9.13)$$

•

Die hier gewählte Definition des Zweistichproben-T-Test-Modells in Designmatrixform ist nicht die einzig mögliche, jedoch diejenige, unter der sich am klarsten die Äquivalenz zum Zweistichproben-T-Test im ALM-freien Kontext erkennen lässt. In Kapitel 10 lernen wir eine alternative Parameterisierung auch des Zweistichproben-T-Test-Modells kennen.

Wie schon beim Szenario des Einstichproben-T-Tests ergibt sich die Äquivalenz der in Definition 9.2 formulierten Modellformen mit den Ergebnissen in Kapitel 5. Die Simulation von Daten basierend auf dem Zweistichproben-T-Test-Modell ist, bis auf die Definition von Designmatrix und Betaparametervektor, mit den bisher bekannten Simulationen von ALM-Spezialfällen identisch, wie folgender **R**-Code demonstriert.

```
# Modellformulierung
library(MASS)                                # multivariate Normalverteilung
n_1 = 12                                     # Anzahl Datenpunkte Gruppe 1
n_2 = 12                                     # Anzahl Datenpunkte Gruppe 2
n = n_1 + n_2                               # Gesamtanzahl Datenpunkte
p = 2                                        # Anzahl Betaparameter
X = matrix(c(rep(1,n_1), rep(0,n_1),      # n x p 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 = 10                                 # wahrer, aber unbekannter Varianzparameter

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

## Modellschätzung

Die beiden Betaparameterkomponenten des Zweistichproben-T-Test-Modells in Designmatrixform werden wenig überraschend durch die entsprechenden Gruppenstichprobenmittel geschätzt. Für den Varianzparameterschätzer ergibt sich die sogenannte *gebündelte Stichprobenvarianz*. Dies sind die beiden Kernaussagen folgenden Theorems.

**Theorem 9.3** (Parameterschätzung im Zweistichproben-T-Test-Modell). *Gegeben sei die Designmatrixform des Zweistichproben-T-Test-Modells. Dann ergeben sich für den Betaparameterschätzer*

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \quad (9.14)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \quad (9.15)$$

◦

*Beweis.* Für  $i = 1, 2$  sei  $y_i := (y_{i1}, \dots, y_{in_i})^T$ . Dann ergibt sich für den Betaparameterschätzer

$$\begin{aligned}
\hat{\beta} &= (X^T X)^{-1} X^T y \\
&= \left( \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
&= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\
&= \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\
&=: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}.
\end{aligned} \tag{9.16}$$

Gleichsam ergibt sich für den Varianzparameterschätzer mit  $n = n_1 + n_2$  und  $p = 2$

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\
&= \frac{1}{n_1 + n_2 - 2} \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right)^T \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right) \\
&= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix} \\
&= \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \\
&=: s_{12}^2.
\end{aligned} \tag{9.17}$$

□

Man beachte, dass sich die Stichprobenvarianz  $s_y^2$  der Komponenten von  $y$  im Allgemeinen von der gebündelten Stichprobenvarianz  $s_{12}^2$  unterscheidet. Dies ist nicht zuletzt dadurch bedingt, dass die Stichprobenvarianz basierend auf dem Gesamtstichprobenmittel  $\bar{y}$ , die gebündelte Stichprobenvarianz dagegen basierend auf den gruppenspezifischen Stichprobenmitteln  $\bar{y}_1$  und  $\bar{y}_2$  ermittelt wird. Wir wollen das Konzept der gebündelten Stichprobenvarianz hier aber nicht weiter vertiefen.

## Modellevaluation

Basierend auf Theorem 7.1 formulieren wir nun die T-Teststatistik für das in Definition 9.2 in Designmatrixform definierte Zweistichproben-T-Test-Modell und geben ihre frequentistische Verteilung an.

**Theorem 9.4** (T-Teststatistik des Zweistichproben-T-Tests). *Gegeben sei die Designmatrixform des Zweistichproben-T-Test-Modells. Dann ergibt sich für die T-Teststatistik mit*

$$c := (1, -1)^T \text{ und } c^T \beta_0 =: \mu_0, \tag{9.18}$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (9.19)$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (9.20)$$

◦

*Beweis.* Mit dem Theorem zur Verteilung der T-Statistik (siehe Theorem 7.1) gilt zunächst für die Zähler von  $T$  und  $\delta$ , dass

$$c^T \hat{\beta} - c^T \beta_0 = (1 \quad -1) \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} - \mu_0 = \bar{y}_1 - \bar{y}_2 - \mu_0 \quad (9.21)$$

und

$$c^T \beta - c^T \beta_0 = (1 \quad -1) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0. \quad (9.22)$$

Weiterhin gilt für die Nenner von  $T$  und  $\delta$ , dass

$$c^T (X^T X)^{-1} c = (1 \quad -1) \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1/n_1 & -1/n_2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2}. \quad (9.23)$$

Außerdem gilt

$$\left( \frac{1}{n_1} + \frac{1}{n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_1 + n_2}{n_1 n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_1 n_2}{n_1 + n_2} \right)^{\frac{1}{2}}. \quad (9.24)$$

Zusammengenommen folgt dann, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \text{ und } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (9.25)$$

□

Die Formen der T-Teststatistik und ihre Verteilung im Zweistichproben-T-Test-Modell in Designmatrixform sind also wiederum natürlicherweise mit den entsprechenden Formen im ALM-freien Kontext identisch. Die entsprechende Theorie zur Kontrolle des Testumfangs bei Zweistichproben-T-Tests sowie der Gebrauch der Testgütefunktion zur Evaluation der Testtrennschärfe (Power) folgt also analog.

## Anwendungsbeispiel

Folgender **R**-Code demonstriert die Evaluation von 95%-Konfidenzintervallen für die Erwartungswertparameter  $\mu_1$  und  $\mu_2$  sowie die Durchführung eines zweiseitigen Zweistichproben-T-Tests mit der Nullhypothese

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2 \mid \mu_1 = \mu_2 \right\} \quad (9.26)$$

und dem Signifikanzlevel  $\alpha_0 := 0.05$  für das oben skizzierte Anwendungsbeispiel.

```

# Dateneinlesen
D = read.csv("data/09_T-Tests.csv") # Dataframe
y_1 = D$BDI[D$COND == "F2F"] # BDI-Differenzwerte in der F2F-Gruppe
y_2 = D$BDI[D$COND == "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_2), # Zweistichproben-T-Test-Designmatrix
            rep(0, n_1), rep(1, n_2)),
          nrow = n)

# 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

# Konfidenzintervall
delta = 0.95 # Konfidenzbedingung
t_delta = qt((1+delta)/2, n-1) # \Psi^{-1}((1+\delta)/2, n-1)
lambda = diag(solve(t(X) %*% X)) # \lambda_j Werte
kappa = matrix(rep(NA, p*2), nrow = p) # \beta_j Konfidenzintervallarray
for(j in 1:p){ # Iteration über \beta_j
  kappa[j,1] = beta_hat[j] - sqrt(sigsqr_hat*lambda[j])*t_delta # untere KI-Grenze
  kappa[j,2] = beta_hat[j] + sqrt(sigsqr_hat*lambda[j])*t_delta # obere KI-Grenze
}

# Hypothesentest
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){phi = 1} else {phi = 0} # Test 1_{|T(X)| >= k_alpha_0}
pval = 2*(1-pt(abs(t), n_1+n_2-2)) # p-Wert

```

```

Betaparameterschätzer : 5.17 5.75
95%-Konfidenzintervalle : 2.6 3.18 7.74 8.32
Varianzparameterschätzer : 18.54
Signifikanzniveau : 0.05
Kritischer Wert : 2.07
Zweistichproben-T-Teststatistik : -0.33
Testergebnis : 0
p-Wert : 0.74

```

Die Nullhypothese wird in diesem Fall also nicht verworfen, da der absolute Betrag des Werts der T-Statistik von  $T = -0.33$  den kritischen Wert von  $k_{0.05} = 2.07$  nicht übersteigt. Inferenzstatistisch besteht also keine Evidenz dafür, dass sich der wahre, aber unbekannte Erwartungswertparameter im F2F-Therapiesetting vom wahren, aber unbekanntem Erwartungswertparameter im ONL-Therapiesetting unterscheidet. Die 95%-Konfidenzintervalle für die wahren, aber unbekanntem Erwartungswertparameter  $\mu_1$  und  $\mu_2$  sind  $[2.60, 7.74]$  bzw.  $[3.18, 8.32]$ .

### 9.3 Literaturhinweise

Obwohl die frequentistische Literatur der ersten Hälfte des 20. Jahrhunderts von der Äquivalenz regressions- und varianzanalytischer linearer Modelle durchdrungen ist, fällt es schwer, eine definite Quelle anzugeben, die hinsichtlich der Beschreibung von T-Tests als Spezialfälle des ALM Priorität hätte. Es sei hier deshalb eher allgemein auf Fisher (1925c) und Fisher (1935) verwiesen.

### 9.4 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- $\alpha_0$ -Einstichproben-T-Test wieder.
12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- $\alpha_0$ -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 Einstichproben-T-Tests bei fester Stichprobengröße.
16. Skizzieren Sie die Powerfunktion des Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.
17. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.
18. Geben Sie die Definition des Zweistichproben-T-Test-Modells wieder.
19. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test-Modell wieder.
20. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.
21. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.
22. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests (mit ungerichteter Hypothese) wieder.
23. Skizzieren Sie die Testgütefunktion des zweiseitigen Zweistichproben-T-Tests.
24. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Test wieder.
25. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Tests.
26. Geben Sie die Definition des p-Wertes für einen zweiseitigen Zweistichproben-T-Test wieder.
27. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Zweistichproben-T-Tests ab?
28. Skizzieren Sie die Powerfunktion des Zweistichproben-T-Tests bei fester Stichprobengröße.
29. Skizzieren Sie die Powerfunktion des Zweistichproben-T-Tests bei festem Nichtzentralitätsparameter.

# 10 Einfaktorielle Varianzanalyse

## 10.1 Anwendungsszenario

Das Anwendungsszenario einer einfaktoriellen Varianzanalyse ist durch das Vorliegen von  $n$  univariaten Datenpunkten von zwei oder mehr Gruppen randomisierter experimenteller Einheiten gekennzeichnet, die sich hinsichtlich der Level eines experimentellen Faktors unterscheiden. Ist die Anzahl an Datenpunkten in jeder Gruppe gleich, so spricht man auch von einem *balancierten einfaktoriellen Varianzanalysedesign*. Von den Datenpunkten der  $i$ ten Gruppe bzw. des  $i$ ten Faktorlevels wird dabei angenommen, dass sie Realisierungen von jeweils  $n_i$  unabhängigen und identisch normalverteilten Zufallsvariablen sind, deren wahre, aber unbekanntes Erwartungswertparameter sich potentiell über die Gruppen hinweg unterscheiden und deren wahrer, aber unbekannter Varianzparameter über Gruppen hinweg identisch ist. In diesen Grundannahmen handelt es sich bei dem Szenario der einfaktoriellen Varianzanalyse also um eine direkte Generalisierung des Einstichproben- und Zweistichproben-T-Test-Szenarios zu (potentiell) mehr als zwei Gruppen. Umgekehrt können die Einstichproben- und Zweistichproben-T-Test-Szenarien natürlich auch als einfaktorielle Varianzanalyseszenarien betrachtet werden, bei denen der experimentelle Faktor (nur) ein oder zwei Level aufweist. Schließlich wird wie im Falle der T-Test-Szenarien meist vorausgesetzt, dass ein Interesse an einem inferentiellen Vergleich der wahren, aber unbekanntes Faktorlevel-spezifischen Erwartungswertparameter besteht.

## 10.2 Anwendungsbeispiel

Als konkretes Anwendungsbeispiel betrachten wir die Analyse von BDI-Differenzwerten in einer Pre-Post-Interventionsstudie von drei Gruppen von jeweils 12 Patient:innen, die unterschiedliche Therapiesettings (Face-to-Face und Online) bzw. eine Wartelistenkontrollbedingung durchlaufen haben, wie in Tabelle 10.1 exemplarisch dargestellt. Die erste Spalte der Tabelle (COND) listet das Patient:innen-spezifische Therapiesetting (F2F: face-to-face, ONL: online, WLC: waitlist control) für jeweils drei Patient:innen jeder Studiengruppe auf. Die zweite Spalte der Tabelle (dBDI) listet die entsprechenden Patient:innenspezifischen BDI-Differenzwerte. Positive Werte entsprechen hier wieder einer Abnahme der Depressionssymptomatik, negative Werte einer Zunahme der Depressionssymptomatik.

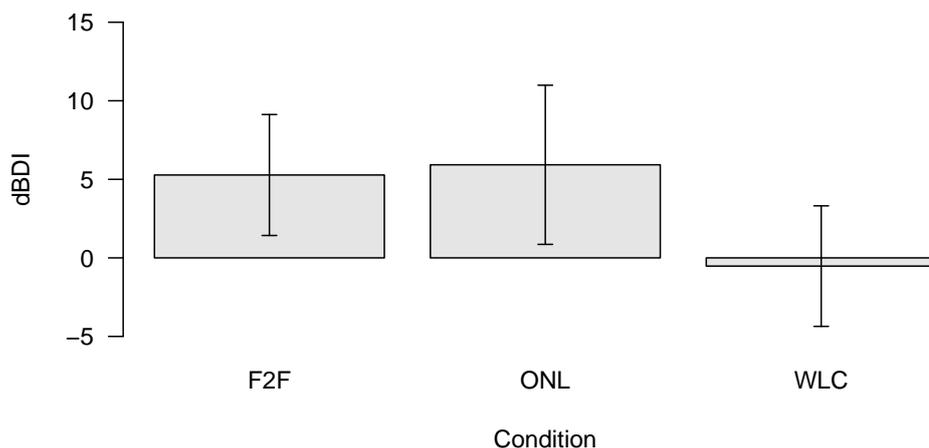
**Tabelle 10.1** Exemplarische BDI-Differenzwerte des Beispieldatensatzes

	COND	dBDI
1	F2F	9
2	F2F	7
3	F2F	10
4	F2F	11
13	ONL	2
14	ONL	7
15	ONL	9

**Tabelle 10.1** Exemplarische BDI-Differenzwerte des Beispieldatensatzes

	COND	dBDI
16	ONL	9
17	ONL	8
25	WLC	-1
26	WLC	2
27	WLC	-3
28	WLC	-1
29	WLC	0

Abbildung 10.1 zeigt eine Visualisierung des gesamten Datensatzes. Die Balken repräsentieren die gruppenspezifischen Stichprobenmittelwerte, die zugehörigen Fehlerbalken die gruppenspezifischen Stichprobenstandardabweichungen. Rein numerisch betrachtet ist in der F2F- und der ONL-Gruppe die Veränderung des BDI-Wertes stärker ausgeprägt als in der WLC-Gruppe.

**Abbildung 10.1** Datendarstellung des Anwendungsbeispiels zur Einfaktoriellen Varianzanalyse.

In Tabelle 10.2 fassen wir die Deskriptivstatistiken des Beispieldatensatzes, aufgeschlüsselt nach Therapiebedingungen, zusammen.

**Tabelle 10.2** Deskriptivstatistiken der BDI-Differenzwerte

	n	Max	Min	Median	Mean	Var	Std
F2F	12	13	6	10.0	9.83	3.79	1.95
ONL	12	10	2	7.5	7.00	7.27	2.70
WLC	12	2	-4	-1.0	-1.42	4.45	2.11

### 10.3 Modellformulierung

Wir definieren das Modell der EVA zunächst in Erwartungswertparameterdarstellung. Dabei nutzen wir den Index  $i$ , um die experimentellen Gruppen zu indizieren und den Index

$j$ , um die experimentellen Einheiten innerhalb der Gruppen zu indizieren.

**Definition 10.1** (EVA-Modell in Erwartungswertparameterdarstellung). Für  $i = 1, \dots, p$  und  $j = 1, \dots, n_i$  seien  $y_{ij}$  Zufallsvariablen, die die  $n := \sum_{i=1}^p n_i$  Datenpunkte eines EVA-Szenarios modellieren. Dann hat das EVA-Modell in Erwartungswertparameterdarstellung die strukturelle Form

$$y_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. mit } \mu_i \in \mathbb{R}, \sigma^2 > 0, \quad (10.1)$$

die Datenverteilungsform

$$y_{ij} \sim N(\mu_i, \sigma^2) \text{ u.v. mit } \mu_i \in \mathbb{R}, \sigma^2 > 0 \quad (10.2)$$

und für den  $n$ -dimensionalen Datenvektor definiert als

$$y := \left( y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2}, \dots, y_{p1}, \dots, y_{pn_p} \right)^T \quad (10.3)$$

die Designmatrixform

$$y = X\beta + \varepsilon \quad (10.4)$$

mit

$$X := \begin{pmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_p} & 0_{n_p} & \cdots & 1_{n_p} \end{pmatrix}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \in \mathbb{R}^p, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (10.5)$$

Wenn  $n_i := m$  für alle  $i = 1, \dots, p$ , heißt das EVA-Szenario *balanciert*.

•

Der Vergleich mit der Definition des Modells des Zweistichproben-T-Tests in Definition 9.2 zeigt, dass es sich bei dieser Formulierung des EVA-Modells um die direkte Generalisierung des Zweistichproben-T-Test-Modells von  $p = 2$  auf ein beliebiges  $p \in \mathbb{N}$  handelt.

## Motivation der Effektdarstellung

Das EVA-Modell in Erwartungswertparameterdarstellung ist ein valides Modell, auf dessen Grundlage sowohl Parameterschätzer als auch Parameter- und Modellinferenz für das EVA-Szenario entwickelt werden können (vgl. Georgii (2009)). Im Sinne der Konsistenz mit den Modellen der mehrfaktoriellen Varianzanalyse bietet sich jedoch eine Reparametrisierung des Betaparametervektors an. Kern dieser Reparametrisierung 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:

$$\mu_i := \mu_0 + \alpha_i \text{ für } i = 1, \dots, p. \quad (10.6)$$

Dabei modelliert  $\alpha_i$  die Differenz zwischen dem  $i$ ten Erwartungswertparameter  $\mu_i$  und dem gruppenübergreifenden Erwartungswertparameter  $\mu_0$ :

$$\alpha_i = \mu_i - \mu_0 \text{ für } i = 1, \dots, p. \quad (10.7)$$

Allerdings hat die in dieser Form vorgenommene Reparametrisierung einen entscheidenden Nachteil: Es werden  $p$  Erwartungswertparameter  $\mu_i, i = 1, \dots, p$  durch die  $p + 1$  Parameter  $\mu_0$  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 in der Form von Gleichung 10.6 überparametrisiert ist.

Datenanalytisch hat die Überparametrisierung 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 einer Nebenbedingung an:

$$\alpha_1 := 0. \quad (10.8)$$

Es wird also ein Effektparameter von vornherein als identisch Null angenommen. Für die gruppenspezifischen Erwartungswertparameter ergibt sich damit:

$$\begin{aligned} \mu_1 &:= \mu_0 \\ \mu_i &:= \mu_0 + \alpha_i \text{ für } i = 2, \dots, p. \end{aligned} \quad (10.9)$$

Hierbei wird die erste Gruppe nun als Referenzgruppe bezeichnet und die  $\alpha_i$  modellieren die Differenz zwischen dem Erwartungswertparameter der  $i$ ten Gruppe und dem Erwartungswertparameter der ersten Gruppe:

$$\alpha_i = \mu_i - \mu_0 = \mu_i - \mu_1 \text{ für } i = 1, \dots, p. \quad (10.10)$$

$\mu_0$  ist, unter der Nebenbedingung  $\alpha_1 := 0$ , also kein gruppenübergreifender Erwartungswertparameter mehr, sondern identisch mit dem Erwartungswertparameter der ersten Gruppe. Welche tatsächliche experimentelle Gruppe dabei als "erste Gruppe" definiert wird, ist datenanalytisch unerheblich. Datenanalytisch entscheidend dagegegen ist, dass der entsprechenden Erwartungswertparameterschätzer  $\hat{\mu}_0$  korrekt als Erwartungswertparameterschätzer der Referenzgruppe und die  $\hat{\alpha}_i$  für  $i = 2, \dots, p$  korrekt als geschätzte Erwartungswertparameterdifferenzen zwischen der Referenzgruppe und der  $i$ ten Gruppe verstanden werden. Wir formalisieren das oben Gesagte in folgendem Theorem.

**Theorem 10.1** (EVA-Modell in Effektdarstellung mit Referenzgruppe). *Gegeben sei das EVA-Modell in Erwartungswertparameterdarstellung. Dann können die Zufallsvariablen, die die Datenpunkte des EVA-Szenarios modellieren, äquivalent in der strukturellen Form*

$$\begin{aligned} y_{1j} &= \mu_0 + \varepsilon_{1j} \quad \text{mit } \varepsilon_{1j} \sim N(0, \sigma^2) \text{ u.i.v. für } j = 1, \dots, n_1 \\ y_{ij} &= \mu_0 + \alpha_i + \varepsilon_{ij} \quad \text{mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 2, \dots, p, j = 1, \dots, n_i \end{aligned} \quad (10.11)$$

mit  $\alpha_i := \mu_i - \mu_1$  für  $i = 2, \dots, p$  und in der entsprechenden Datenverteilungsform

$$\begin{aligned} y_{1j} &\sim N(\mu_0, \sigma^2) && \text{u.i.v. für } j = 1, \dots, n_1 \text{ mit } \mu_0 \in \mathbb{R}, \sigma^2 > 0 \\ y_{ij} &\sim N(\mu_0 + \alpha_i, \sigma^2) && \text{u.v. für } i = 2, \dots, p, j = 1, \dots, n_i \text{ mit } \alpha_i \in \mathbb{R}, \sigma^2 > 0 \end{aligned} \quad (10.12)$$

geschrieben werden. Wir nennen (10.11) und (10.12) die strukturelle und die Datenverteilungsform des EVA-Modells in Effektdarstellung mit Referenzgruppe.

◦

*Beweis.* Es gilt

$$\mu_i = \mu_0 + \mu_i - \mu_0. \quad (10.13)$$

Die Parameterisierungen mit  $\mu_i$  und mit  $\mu_0 + \mu_i - \mu_0$  sind also gleich und damit äquivalent. Dann folgt aber auch

$$\mu_i = \mu_0 + (\mu_i - \mu_0) =: \mu_0 + \alpha_i \text{ für } i = 1, \dots, p. \quad (10.14)$$

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. □

Basierend auf Theorem 10.1 definieren wir nun das Modell der EVA in Effektdarstellung.

**Definition 10.2** (EVA-Modell in Effektdarstellung mit Referenzgruppe). Gegeben sei die strukturelle Form des EVA-Modells in Effektdarstellung mit Referenzgruppe. Dann hat dieses Modell die Designmatrixform

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n), n := \sum_{i=1}^p n_i \quad (10.15)$$

$$y := \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix} \in \mathbb{R}^n, \quad X := \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{n \times p}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \text{ und } \sigma^2 > 0.$$

•

## Beispiel

Um die Unterschiede und Gemeinsamkeiten zwischen der Erwartungswertparameterdarstellung und der Effektparameterdarstellung des EVA-Modells in ihrer Designmatrixform zu verdeutlichen, betrachten wir ein Beispielszenario mit  $p = 3$  und  $n_i := 4$  für  $i = 1, \dots, p$ , also  $n = 12$ . Für die Erwartungswertparameterdarstellung gilt dann

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{12}, \sigma^2 I_{12}) \quad (10.16)$$

mit

$$y := \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{12 \times 3}, \quad \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (10.17)$$

Für die Effektparameterdarstellung dagegen gilt

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{12}, \sigma^2 I_{12}) \quad (10.18)$$

mit

$$y := \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{12 \times 3}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (10.19)$$

Folgender **R**-Code demonstriert die Realisierung von Daten in einem EVA-Szenario auf Grundlage dieser Effektparameterdarstellung.

```
# Modellformulierung
library(MASS)
m = 4
p = 3
n = p*m
Xt = cbind(
  matrix(1, nrow = n, ncol = 1),
  kronecker(diag(p), matrix(1, nrow = m, ncol = 1)))
# multivariate Normalverteilung
# Anzahl von Datenpunkten pro Gruppe
# Anzahl von Gruppen
# Gesamtanzahl Datenpunkte
# n x p Designmatrix
```

```

X      = Xt[,-2]
I_n    = diag(n)
beta   = matrix(c(10,-3,-12), nrow = p)
sigsqr = 14
y      = mvrnorm(1, X %*% beta, sigsqr*I_n)
print(X)
# n x n Einheitsmatrix
# \beta = (\mu_0, \alpha_2, \alpha_3, \alpha_4)
# \sigma^2
# eine Realisierung des n-dimensionalen ZVs y

```

```

      [,1] [,2] [,3]
[1,]  1   0   0
[2,]  1   0   0
[3,]  1   0   0
[4,]  1   0   0
[5,]  1   1   0
[6,]  1   1   0
[7,]  1   1   0
[8,]  1   1   0
[9,]  1   0   1
[10,] 1   0   1
[11,] 1   0   1
[12,] 1   0   1

```

## 10.4 Modellschätzung

Wir betrachten nun die Betaparameterschätzung in der Effektparameterdarstellung des EVA-Modells mit Referenzgruppe. Entsprechend der Interpretation der Betaparameterkomponenten werden dabei  $\mu_0$  durch das Stichprobenmittel der Referenzgruppe und die  $\alpha_2, \dots, \alpha_p$  durch die Differenzen des jeweiligen Gruppenstichprobenmittels und des Referenzgruppenstichprobenmittels geschätzt. Auf die Schätzung des Varianzparameters, der sich wie im Zweistichproben-T-Test-Modell zu einer *gebündelten Stichprobenvarianz* ergibt, wollen wir hier nicht weiter eingehen.

**Theorem 10.2** (Betaparameterschätzung im EVA-Modell). *Gegeben sei die Designmatrixform der EVA in Effektdarstellung mit Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer*

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_p - \bar{y}_1 \end{pmatrix}, \quad (10.20)$$

wobei

$$\bar{y}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad (10.21)$$

das Stichprobenmittel der *i*ten Gruppe bezeichnet.

◦

*Beweis.* Wir halten zunächst fest, dass

$$\begin{aligned}
 X^T X &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & 0 \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} n & n_2 & n_3 & \dots & n_p \\ n_2 & n_2 & 0 & \dots & 0 \\ n_3 & 0 & n_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_p & 0 & 0 & \dots & n_p \end{pmatrix}.
 \end{aligned}$$

Die Inverse von  $X^T X$  ist

$$(X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & \dots & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \dots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_1} & \frac{1}{n_1} & \dots & \frac{n_1+n_p}{n_1 n_p} \end{pmatrix}. \tag{10.22}$$

So gilt zum Beispiel für  $p = 3$ , dass

$$X^T X = \begin{pmatrix} n & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{pmatrix} \text{ und } (X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{1}{n_1} & \frac{n_1+n_3}{n_1 n_3} \end{pmatrix}. \tag{10.23}$$

Wir halten weiterhin fest, dass

$$\begin{aligned}
 X^T y &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \\ \sum_{j=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} y_{pj} \end{pmatrix}.
 \end{aligned} \tag{10.24}$$

Es ergibt sich also

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & \dots & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \dots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_1} & \frac{1}{n_1} & \dots & \frac{n_1+n_p}{n_1 n_p} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \\ \sum_{j=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} y_{pj} \end{pmatrix}. \quad (10.25)$$

Für die erste Komponente von  $\hat{\beta}$  ergibt sich damit

$$\begin{aligned} \hat{\beta}_0 &= \frac{1}{n_1} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} - \frac{1}{n_1} \sum_{j=1}^{n_2} y_{2j} - \dots - \frac{1}{n_1} \sum_{j=1}^{n_p} y_{pj} \\ &= \frac{1}{n_1} \left( \left( \sum_{j=1}^{n_1} y_{1j} + \sum_{j=1}^{n_2} y_{2j} + \dots + \sum_{j=1}^{n_p} y_{pj} \right) - \sum_{j=1}^{n_2} y_{2j} - \dots - \sum_{j=1}^{n_p} y_{pj} \right) \\ &= \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ &= \bar{y}_1. \end{aligned} \quad (10.26)$$

Für die zweite Komponente von  $\hat{\beta}$  und analog für alle weiteren ergibt sich

$$\begin{aligned} \hat{\beta}_2 &= -\frac{1}{n_1} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} + \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} + \frac{1}{n_1} \sum_{j=1}^{n_3} y_{3j} + \dots + \frac{1}{n_1} \sum_{j=1}^{n_p} y_{pj} \\ &= \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \left( \left( \sum_{j=1}^{n_1} y_{1j} + \sum_{j=1}^{n_2} y_{2j} + \dots + \sum_{j=1}^{n_p} y_{pj} \right) - \sum_{j=1}^{n_3} y_{3j} - \dots - \sum_{j=1}^{n_p} y_{pj} \right) \\ &= \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} - \frac{1}{n_1} \sum_{j=1}^{n_2} y_{2j} \\ &= \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{n_2}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ &= \frac{n_1}{n_1 n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ &= \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} = \bar{y}_2 - \bar{y}_1. \end{aligned} \quad (10.27)$$

□

Folgender **R**-Code implementiert die Parameterschätzung des EVA-Modells für den Beispieldatensatz. Neben der mithilfe des Betaparameterschätzers gewonnenen Schätzwerte für den Referenzgruppen- und die Effektparameter evaluiert der Code auch die Stichprobenmittel(differenzen)-basierten Schätzer aus Theorem 10.2. Weiterhin wird neben dem Varianzparameterschätzer auch die gebündelte Stichprobenvarianz sowie die Gesamtstichprobenvarianz ausgegeben, die sich deutlich voneinander unterscheiden.

```
# Datenmanagement
D = read.csv("data/10_Einfaktorielle_Varianzanalyse.csv") # Dataframe
y = D$dBDDI # Datenvektor
y_1 = D$dBDDI[D$COND == "F2F"] # Daten F2F
y_2 = D$dBDDI[D$COND == "ONL"] # Daten ONL
y_3 = D$dBDDI[D$COND == "WLC"] # Daten WLC

# Modellformulierung
p = 3 # drei Gruppen
m = length(y_1) # balanciertes Design
```

```

n      = p*m                                # Datenvektordimension
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) +                # gebündelte Stichprobenvarianz
  (m-1)*var(y_2) +
  (m-1)*var(y_3))/(m+m-m-p)

hat{beta} : 9.83 -2.83 -11.25
bar{y}_1, bar{y}_2, bar{y}_3 : 9.83 7 -1.42
bar{y}_1, bar{y}_2 - bar{y}_1, bar{y}_3 - bar{y}_1 : 9.83 -2.83 -11.25
hat{sigsqr} : 5.17
s_123^2 : 5.17
s_y^2 : 28.35

```

## 10.5 Modellevaluation

Prinzipiell sind alle Parameterschätzwerte in einem EVA-Modell von Interesse und können mithilfe von T-Statistiken im Sinne von Konfidenzintervallen oder Hypothesentests evaluiert werden. Traditionell steht im EVA-Szenario allerdings häufig die Evaluation der Nullhypothese, dass die wahren, aber unbekannt, Erwartungswertparameter aller Gruppen identisch sind, im Vordergrund. Vor dem Hintergrund der EVA-Effektdarstellung mit Referenzgruppe entspricht dies der Nullhypothese, dass die Effektparameter alle gleich Null sind, formal ausgedrückt entspricht die Nullhypothese der Menge

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \mid \alpha_i = 0 \text{ für } i = 2, \dots, p \right\} = \mathbb{R} \times \{0_{p-1}\} \quad (10.28)$$

und die Alternativhypothese entspricht der Menge

$$\Theta_1 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \mid \alpha_i \neq 0 \text{ für mindestens ein } i = 2, \dots, p \right\} = \mathbb{R}^p \setminus \Theta_0. \quad (10.29)$$

Zur Evaluation der Nullhypothese wird dabei im Allgemeinen eine F-Statistik genutzt. Im Folgenden wollen wir zunächst diese F-Statistik anhand einer Quadratsummenzerlegung der Datenvariabilität in einem EVA-Szenario entwickeln. In diesem Kontext führen wir mit  $\eta^2$  (*Eta-Quadrat*) auch ein zu dem aus Kapitel 2 bekannten Bestimmtheitsmaß  $R^2$  analoges Effektstärkenmaß ein. Ausgestattet mit der speziellen Form der F-Statistik für das EVA-Modell diskutieren wir dann den traditionellen Test der EVA-Nullhypothese.

### 10.5.1 Quadratsummenzerlegung und Bestimmtheitsmaß $\eta^2$

Die Variabilität der Daten eines EVA-Szenarios lässt sich im Sinne einer Quadratsummenzerlegung schreiben, wie in folgendem Theorem dargestellt wird.

**Theorem 10.3** (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

$$SQT := \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \quad \text{die total sum of squares}$$

$$SQB := \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2 \quad \text{die "between" sum of squares}$$

$$SQW := \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad \text{die "within" sum of squares} \quad .$$

Dann gilt

$$SQT = SQB + SQW. \tag{10.30}$$

◦

*Beweis.* Es gilt

$$\begin{aligned} SQT &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i + \bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} ((y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}))^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} ((y_{ij} - \bar{y}_i)^2 + 2(y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}) + (\bar{y}_i - \bar{y})^2) \\ &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{j=1}^{n_i} 2(y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}) + \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) + n_i (\bar{y}_i - \bar{y})^2 \right) \\ &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \sum_{j=1}^{n_i} \left( y_{ij} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \right) + n_i (\bar{y}_i - \bar{y})^2 \right) \end{aligned} \tag{10.31}$$

und weiter

$$\begin{aligned}
 \text{SQT} &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \left( \sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \right) \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \\
 &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \left( \sum_{j=1}^{n_i} y_{ij} - \frac{n_i}{n_i} \sum_{j=1}^{n_i} y_{ij} \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \\
 &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y}) \left( \sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} y_{ij} \right) + n_i(\bar{y}_i - \bar{y})^2 \right) \\
 &= \sum_{i=1}^p \left( \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + n_i(\bar{y}_i - \bar{y})^2 \right) \\
 &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^p n_i(\bar{y}_i - \bar{y})^2
 \end{aligned}$$

und damit

$$\text{SQT} = \text{SQB} + \text{SQW}. \tag{10.32}$$

□

Die Quadratsummenzerlegung im Szenario der EVA ist offenbar analog zur Quadratsummenzerlegung bei einer Ausgleichsgerade (vgl. Kapitel 1.1). Dementsprechend wird die Within-Sum-of-Squares der EVA auch häufig als Residual-Sum-of-Squares bezeichnet. Die Begriffsbildungen von Theorem 10.3 ergeben sich intuitiv entsprechend analog:

- SQT repräsentiert die Gesamtvariabilität der Daten  $y_{ij}$  um das Gesamtstichprobenmittel  $\bar{y}$ .
- SQB repräsentiert die anhand des jeweiligen Gruppenumfangs  $n_i$  gewichtete Variabilität der Gruppenstichprobenmittel um das Gesamtstichprobenmittel. Große Werte von SQB implizieren also eine große Abhängigkeit der Gruppenstichprobenmittel von dem jeweils betrachteten Faktorlevel  $i$ , kleine Werte von SQB dagegen eine geringe Abhängigkeit der Gruppenstichprobenmittel von dem jeweils betrachteten Faktorlevel. In diesem Sinne repräsentiert SQB also die durch die Betrachtung des jeweiligen Faktorlevels erklärte Datenvariabilität und ist analog zur erklärten Quadratsumme (SQE) der Quadratsummenzerlegung bei einer Ausgleichsgerade.
- SQW repräsentiert die über Faktorlevel summierte Datenvariabilität, die nach Erklärung der Datenvariabilität in der  $i$ ten Gruppe durch ihr jeweiliges Stichprobenmittel verbleibt. Damit ist die SQW analog zur residuellen Quadratsumme der Quadratsummenzerlegung bei Ausgleichsgerade.

In der Zusammenschau quantifiziert SQB also die Stärke der Unterschiede zwischen den Faktorleveln. Das Effektstärkenmaß  $\eta^2$  und die F-Statistik der EVA setzen die SQB nun jeweils in ein anderes Verhältnis:  $\eta^2$  vergleicht die SQB mit der SQT und betrachtet damit den Anteil der Variabilität zwischen den Faktorleveln an der Gesamtvariabilität der Daten. Die F-Statistik dagegen vergleicht die SQB mit der SQW und setzt damit den Einfluß der Faktorenlevel mit der nicht erklärten Datenvariabilität nach Subtraktion dieses Einflusses ins Verhältnis. Wir betrachten an dieser Stelle zunächst das Effektstärkemaß  $\eta^2$  mithilfe folgender Definition.

**Definition 10.3** (Effektstärkenmaß  $\eta^2$ ). 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  $\eta^2$  definiert als

$$\eta^2 := \frac{SQB}{SQT} \quad (10.33)$$

•

Der Vergleich mit Definition 2.4 zeigt, dass  $\eta^2$  analog zum Bestimmtheitsmaß  $R^2$  definiert ist. Wie oben beschrieben gibt  $\eta^2$  den Anteil der Datenvariabilität zwischen den Faktorleveln an der Gesamtdatenvariabilität wieder. Schließlich folgt mit

$$SQT = SQB + SQW \quad (10.34)$$

analog zu  $R^2$  sofort, dass für  $SQT \neq 0$  gilt, dass  $\eta^2 \in [0, 1]$ , weil einerseits

$$SQB = 0 \Rightarrow SQT = SQW \text{ und } \eta^2 = 0 \quad (10.35)$$

und andererseits

$$SQW = 0 \Rightarrow SQT = SQB \text{ und } \eta^2 = 1. \quad (10.36)$$

### 10.5.2 F-Teststatistik

Wir wollen nun zeigen, dass für die Designmatrixform der Effektdarstellung mit Referenzgruppe des EVA-Modells die F-Statistik bei der Partitionierung  $p_0 := 1$  und  $p_1 := p - 1$  (vgl. Kapitel 8) als Verhältnis skalierten Versionen der SQB und SQW geschrieben werden kann. Dabei impliziert  $p_0 := 1$  hier insbesondere, dass das betrachtete reduzierte EVA-Modell die Designmatrix  $X_0 := 1_n$  und den Betaparameter  $\beta := \mu_0$  hat, und damit insbesondere auch, dass das reduzierte EVA-Modell keine Effektparameter hat. Die dabei relevanten Skalierungsfaktoren beziehen die SQB und SQW jeweils auf die Anzahl der Effektparameter bzw. die Differenz aus Gesamtanzahl der Datenpunkte bzw. Anzahl der Betaparameter. Es gilt folgendes Theorem.

**Theorem 10.4** (F-Statistik der einfaktoriellen Varianzanalyse). *Es sei*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (10.37)$$

*die Designmatrixform der Effektdarstellung des EVA-Modells mit Referenzgruppe und im Sinne der Definition der F-Statistik sei dieses Modell partitioniert mit  $p_0 := 1$  und  $p_1 := p - 1$ . Weiterhin seien*

$$MSB := \frac{SQB}{p-1} \quad \text{die mean between sum of squares und}$$

$$MSW := \frac{SQW}{n-p} \quad \text{die mean within sum of squares.}$$

*Dann gilt*

$$F = \frac{MSB}{MSW}. \quad (10.38)$$

◦

*Beweis.* Wir halten zunächst fest, dass für den Betaparameterschätzer des reduzierten Modells gilt, dass

$$\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y = (1_n^T 1_n)^{-1} 1_n^T y = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} = \bar{y}. \quad (10.39)$$

Weiterhin ergibt sich

$$\hat{\varepsilon}_0^T \hat{\varepsilon}_0 = (y - X_0 \hat{\beta}_0)^T (y - X_0 \hat{\beta}_0) = (y - 1_n \bar{y})^T (y - 1_n \bar{y}) = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \text{SQT}. \quad (10.40)$$

Der Betaparameterschätzer des vollständigen Modells ergibt sich wie oben gesehen zu

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_m \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \vdots \\ \frac{1}{n_m} \sum_{j=1}^{n_m} y_{mj} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_p - \bar{y}_1 \end{pmatrix}, \quad (10.41)$$

sodass

$$\begin{aligned} \hat{\varepsilon}^T \hat{\varepsilon} &= (y - X \hat{\beta})^T (y - X \hat{\beta}) \\ &= \begin{pmatrix} \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 0 \\ 1 & 1 & & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 1 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \\ \vdots \\ \bar{y}_p - \bar{y}_1 \end{pmatrix} \end{pmatrix}^T (y - X \hat{\beta}) \end{aligned}$$

und weiter

$$\begin{aligned} \hat{\varepsilon}^T \hat{\varepsilon} &= \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{2n_2} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{p1} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \\ \vdots \\ y_{pn_p} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{2n_2} - \bar{y}_1 - \bar{y}_2 + \bar{y}_1 \\ \vdots \\ y_{p1} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \\ \vdots \\ y_{pn_p} - \bar{y}_1 - \bar{y}_p + \bar{y}_1 \end{pmatrix} = \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \\ \vdots \\ y_{p1} - \bar{y}_p \\ \vdots \\ y_{pn_p} - \bar{y}_p \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \\ \vdots \\ y_{p1} - \bar{y}_p \\ \vdots \\ y_{pn_p} - \bar{y}_p \end{pmatrix} \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \\ &= \text{SQW}. \end{aligned}$$

Mit dem Theorem zur Quadratsummenzerlegung bei einfaktorieller Varianzanalyse

$$\text{SQT} = \text{SQB} + \text{SQW} \Leftrightarrow \text{SQB} = \text{SQT} - \text{SQW} \quad (10.42)$$

folgt sofort, dass

$$\begin{aligned} \text{SQB} &= \text{SQT} - \text{SQW} \\ &= \hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}. \end{aligned} \tag{10.43}$$

Dann aber folgt auch direkt, dass

$$\frac{\text{MSB}}{\text{MSW}} = \frac{\frac{\text{SQB}}{p-1}}{\frac{\text{SQW}}{n-p}} = \frac{\frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{p-1}}{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p}} = F \tag{10.44}$$

□

Gegeben die sehr ähnlichen Definitionen des Effektstärkemaßes  $\eta^2$  und der F-Statistik der EVA ist folgendes Resultat naheliegend.

**Theorem 10.5** (Effektstärkenmaß  $\eta^2$  und F-Teststatistik). *Für ein EVA-Szenario mit  $p$  Gruppen und Gesamtdatenpunktzahl  $n$  seien das Effektstärkenmaß  $\eta^2$  und die F-Teststatistik wie oben definiert. Dann gilt*

$$\eta^2 = \frac{F(p-1)}{F(p-1) + (n-p)} \tag{10.45}$$

◦

*Beweis.* Wir halten zunächst fest, dass

$$F = \frac{\text{SQB}}{\text{SQW}} \cdot \frac{n-p}{p-1} \Leftrightarrow \text{SQB} = \frac{p-1}{n-p} \cdot \text{SQW} \cdot F. \tag{10.46}$$

Damit folgt dann

$$\begin{aligned} \eta^2 &= \frac{\text{SQB}}{\text{SQT}} = \frac{\text{SQB}}{\text{SQB} + \text{SQW}} = \frac{\frac{p-1}{n-p} \cdot \text{SQW} \cdot F}{\frac{p-1}{n-p} \cdot \text{SQW} \cdot F + \text{SQW}} \\ &= \frac{\frac{F(p-1)}{n-p} \cdot \text{SQW}}{\frac{F(p-1)}{n-p} \cdot \text{SQW} + \text{SQW}} = \frac{\frac{F(p-1)}{n-p} \cdot \text{SQW}}{\left(\frac{F(p-1)}{n-p} + 1\right) \cdot \text{SQW}} \\ &= \frac{\frac{F(p-1)}{n-p}}{\frac{F(p-1)}{n-p} + \frac{n-p}{n-p}} = \frac{\frac{F(p-1)}{n-p}}{\frac{F(p-1) + (n-p)}{n-p}} = \frac{F(p-1)}{F(p-1) + (n-p)}. \end{aligned} \tag{10.47}$$

□

Intuitiv ist dabei das Verhältnis von  $\eta^2$  und F-Statistik analog zum Verhältnis von Cohen's  $d$  und der T-Statistik bei Einstichproben- und Zweistichproben-T-Tests. Insbesondere ist die gleichzeitige Angabe von  $\eta^2$  und der F-Statistik bei bekannten Gruppengrößen redundant.

### 10.5.3 F-Test der einfaktoriellen Varianzanalyse

Wir wollen nun den Gebrauch der F-Statistik zur Durchführung eines Tests der Nullhypothese (siehe Gleichung 10.28) diskutieren. Wir erinnern daran, dass diese Nullhypothese intuitiv besagt, dass die Erwartungswerte über alle Level des Faktors identisch sind bzw. dass alle Effektparameter gleich Null sind. Das Verwerfen dieser Nullhypothese impliziert also, dass inferentielle Evidenz dahingehend besteht, dass zumindest ein Effektparameter von Null verschieden ist. Allerdings impliziert das Verwerfen der Nullhypothese des F-Tests der einfaktoriellen Varianzanalyse keine Aussage darüber, um welchen Effektparameter es sich dabei genau handelt. Wir definieren den F-Test der einfaktoriellen Varianzanalyse wie folgt.

**Definition 10.4** (F-Test der einfaktoriellen Varianzanalyse). Gegeben sei das Modell der einfaktoriellen Varianzanalyse sowie die zusammengesetzten Null- und Alternativhypothesen

$$H_0 : \alpha_i = 0 \text{ für alle } i = 2, \dots, p \quad (10.48)$$

bzw.

$$H_1 : \alpha_i \neq 0 \text{ für mindestens ein } i = 2, \dots, p . \quad (10.49)$$

Weiterhin sei die F-Teststatistik definiert durch

$$F = \frac{\text{MSB}}{\text{MSW}} \quad (10.50)$$

mit der *mean between sum of squares* MSB und der *mean within sum of squares* MSW (siehe Theorem 10.4). Dann ist der *F-Test für die einfaktorielle Varianzanalyse (EVA-F-Test)* definiert als der kritische Wert-basierte Test

$$\phi(y) := 1_{\{F \geq k\}} := \begin{cases} 1 & F \geq k \\ 0 & F < k \end{cases} . \quad (10.51)$$

•

Der Kontrolle der Typ-I-Fehlerwahrscheinlichkeit liegt die *f*-Verteilung der F-Statistik zugrunde. Dies ist die Kernaussage folgenden Theorems.

**Theorem 10.6** (Testumfangkontrolle des F-Tests der einfaktoriellen Varianzanalyse).  $\phi$  sei der F-Test zur einfaktoriellen Varianzanalyse. Dann ist  $\phi$  ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$ , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \varphi^{-1}(1 - \alpha_0; p - 1, n - p), \quad (10.52)$$

wobei  $\varphi^{-1}(\cdot; p - 1, n - p)$  die inverse KVF der *f*-Verteilung mit Freiheitsgradparametern  $p - 1$  und  $n - p$  ist.

◦

*Beweis.* Die Testgütefunktion des betrachteten Tests im vorliegenden Testszenario ist definiert als

$$q : \mathbb{R}^p \rightarrow [0, 1], \beta \mapsto q_\phi(\beta) := \mathbb{P}_\beta(\phi = 1). \quad (10.53)$$

Wir haben bereits gesehen (siehe Theorem 8.3), dass die F-Teststatistik für  $p_1 = p - 1$  nach einer nicht-zentralen  $f$ -Verteilung verteilt ist,

$$F \sim f(\delta, p - 1, n - p). \quad (10.54)$$

Weiterhin ist der Ablehnungsbereich des hier betrachteten Tests gegeben als  $[k, \infty[$ . Für die funktionale Form der Testgütefunktion ergibt sich also

$$\begin{aligned} \mathbb{P}_\beta(\phi = 1) &= \mathbb{P}_\beta(F \in [k, \infty[) \\ &= \mathbb{P}_\beta(F \geq k) \\ &= 1 - \mathbb{P}_\beta(F \leq k) \\ &= 1 - \varphi(k; \delta, p - 1, n - p), \end{aligned} \quad (10.55)$$

wobei  $\varphi(k; \delta, p - 1, n - p)$  den KVF-Wert der nichtzentralen  $f$ -Verteilung an Stelle  $k$  mit Nichtzentralitätsparameter  $\delta$  sowie Freiheitsgradparametern  $p - 1$  und  $n - p$  bezeichnet (vgl. Kapitel 8).

Damit der betrachtete Test ein Level- $\alpha_0$ -Test ist, muss bekanntlich gelten, dass

$$q_\phi(\beta) \leq \alpha_0 \text{ für alle } \beta \in \Theta_0 \text{ mit } \Theta_0 = \mathbb{R} \times \{0_{p-1}\}. \quad (10.56)$$

Der Nichtzentralitätsparameter ist gegeben durch (siehe Theorem 8.3)

$$\delta = \frac{1}{\sigma^2} c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta. \quad (10.57)$$

Mit dem dazugehörigen  $c$  und unter der Nullhypothese, dass  $\beta \in \Theta_0$ , also

$$c = \begin{pmatrix} 0 \\ 1_{p-1} \end{pmatrix} \in \mathbb{R}^p \text{ und } \beta = \begin{pmatrix} \mu_0 \\ 0_{p-1} \end{pmatrix} \in \mathbb{R}^p, \quad (10.58)$$

folgt dann aber  $\delta = 0$  und somit

$$q_\phi(\beta) = 1 - \varphi(k; p - 1, n - p) \text{ für alle } \beta \in \Theta_0, \quad (10.59)$$

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. Kapitel 8). Der Testumfang des betrachteten Tests ergibt sich per Definition als

$$\alpha = \max_{\beta \in \Theta_0} q_\phi(\beta) = 1 - \varphi(k; p - 1, n - p), \quad (10.60)$$

da  $q_\phi(\beta)$  für  $\beta \in \Theta_0$  nicht von  $\mu_0$  abhängt. Wir müssen also lediglich zeigen, dass die Wahl von  $k_{\alpha_0}$  wie im Theorem garantiert, dass  $\phi$  den Testumfang  $\alpha_0$  hat. Sei also  $k := k_{\alpha_0}$ . Dann gilt für alle  $\beta \in \Theta_0$

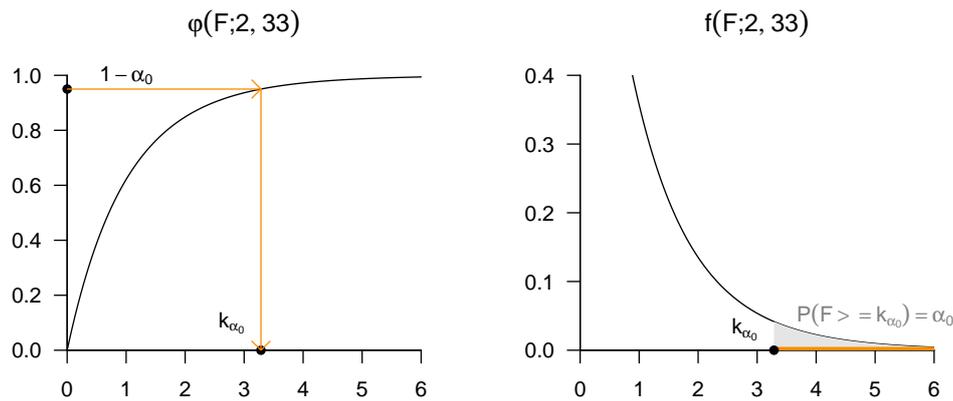
$$q_\phi(\beta) = 1 - \varphi(\varphi^{-1}(1 - \alpha_0; p - 1, n - p); p - 1, n - p) = 1 - (1 - \alpha_0) = \alpha_0 \quad (10.61)$$

und damit ist alles gezeigt. □

In Abbildung 10.2 visualisieren wir die Wahl des kritischen Wertes  $k_{\alpha_0}$  zur Kontrolle des Testumfangs mithilfe der KVF der  $f$ -Verteilung sowie den resultierenden Ablehnungsbereich für  $\alpha_0 := 0.05$ , einen Faktor mit  $p = 3$  Levels und ein balanciertes Design mit Gruppenumfang  $m = 12$ , also  $n = 3 \cdot 12 = 36$  Datenpunkte. In diesem Fall würde die Nullhypothese bei einem Wert der F-Statistik von größer als 3.28 verworfen werden.

### p-Wert

Es sei  $f$  ein vorliegender Wert der  $F$ -Statistik im Kontext einer einfaktoriellen Varianzanalyse. Dann ergibt sich der zu  $f$  gehörige p-Wert anhand folgenden Theorems.



**Abbildung 10.2** Exemplarischer kritischer Wert und Ablehnungsbereich bei einem F-Test der einfaktoriellen Varianzanalyse.

**Theorem 10.7** (p-Wert der F-Statistik bei einfaktorieller Varianzanalyse). *Gegeben sei das in Definition 10.4 spezifizierte Szenario eines F-Tests bei einfaktorieller Varianzanalyse. Dann ergibt sich der zu einem vorliegenden Wert  $f$  der F-Statistik zugehörige p-Wert als*

$$p\text{-Wert} = \mathbb{P}(F \geq f) = 1 - \varphi(f; p - 1, n - p) \tag{10.62}$$

◦

*Beweis.* Nach Definition ist der p-Wert das kleinste Signifikanzlevel  $\alpha_0$ , 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  $\alpha_0$  mit  $f \geq \psi^{-1}(1 - \alpha_0; p - 1, n - p)$  abgelehnt werden. Für ein solches  $\alpha_0$  gilt aber

$$\alpha_0 \geq \mathbb{P}(F \geq f), \tag{10.63}$$

denn

$$\begin{aligned} f &\geq \psi^{-1}(1 - \alpha_0; p - 1, n - p) \\ \Leftrightarrow \psi(f; p - 1, n - p) &\geq \psi(\psi^{-1}(1 - \alpha_0; p - 1, n - p), p - 1, n - p) \\ \Leftrightarrow \psi(f; p - 1, n - p) &\geq 1 - \alpha_0 \\ \Leftrightarrow \mathbb{P}(F \leq f) &\geq 1 - \alpha_0 \\ \Leftrightarrow \alpha_0 &\geq 1 - \mathbb{P}(F \leq f) \\ \Leftrightarrow \alpha_0 &\geq \mathbb{P}(F \geq f) \end{aligned} \tag{10.64}$$

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

$$p\text{-Wert} = \mathbb{P}(F \geq f) = 1 - \varphi(f; p - 1, n - p). \tag{10.65}$$

□

### Praktisches Vorgehen

Aus dem in diesem Abschnitt Gesagten ergibt sich folgendes Vorgehen zur Evaluation eines EVA-Modells mithilfe eines F-Tests: Man nimmt zunächst einmal an, dass ein vorliegender Datensatz von  $p$  Gruppendatensätzen

$$y_{11}, \dots, y_{1n_1}, \quad y_{21}, \dots, y_{2n_2}, \quad \dots, \quad y_{p1}, \dots, y_{pn_p} \quad (10.66)$$

Realisationen von

$$y_{1j} \sim N(\mu_0, \sigma^2) \quad (10.67)$$

und

$$y_{ij} \sim N(\mu_0 + \alpha_i, \sigma^2) \quad (10.68)$$

für  $i = 2, \dots, p$  mit wahren, aber unbekanntem Parametern  $\mu_0, \alpha_i, i = 2, \dots, p$  und  $\sigma^2 > 0$  sind. Weiterhin nimmt man an, dass man entscheiden möchte, ob die Nullhypothese identischer wahrer, aber unbekannter Erwartungswertparameter über Faktorlevel bzw. zu Null identischer wahrer, aber unbekannter Effektparameter eher zutrifft oder eher nicht (vgl. Gleichung 10.28).

Dazu wählt man zunächst ein Signifikanzniveau  $\alpha_0$  und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert  $k_{\alpha_0}$ . Zum Beispiel gilt bei Wahl von  $\alpha_0 := 0.05, p = 3, m = 12$  und somit  $n = 36$ , dass  $k_{\alpha_0} = \varphi^{-1}(1 - 0.05; 2, 33) \approx 3.28$  ist.

Man nutzt dann den vorliegenden Datensatz zur Berechnung der MSB und MSW und bestimmt damit den realisierten Wert der F-Statistik  $f$ . Wenn  $f$  größer gleich  $k_{\alpha_0}$  ist, lehnt man die Nullhypothese ab, andernfalls nicht. Insgesamt garantiert die in diesem Abschnitt entwickelte Theorie dann, dass man im Mittel in höchstens  $(\alpha_0 \cdot 100)$  von 100 Fällen die Nullhypothese fälschlicherweise ablehnt. Schließlich mag man den zu dem Wert  $f$  assoziierten p-Wert als  $1 - \varphi(f; p - 1, n - p)$  bestimmen und in der Dokumentation der Analyse vermerken.

## Anwendungsbeispiel

Folgender **R**-Code implementiert das obige praktische Vorgehen für den Beispieldatensatz.

```
# Datenmanagement
D = read.csv("data/10_Einfaktorielle_Varianzanalyse.csv") # Datenframe
y = D$dBDI # Datenvektor
n = length(y) # Gesamtdatenumfang
p = 3 # Anzahl Gruppen
m = n/p # Anzahl Datenpunkte pro Gruppe

# Modellformulierung
Xt = cbind( # Designmatrix vollständiges Modell
  matrix(1, nrow = n, ncol = 1),
  kronecker(diag(p), matrix(1, nrow = m, ncol = 1)))
X = Xt[,-2]
X_0 = X[,1] # Designmatrix reduziertes Modell

# Evaluation der F-Teststatistik
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer vollständiges Modell
beta_hat_0 = solve(t(X_0) %*% X_0) %*% t(X_0) %*% y # Betaparameterschätzer reduziertes Modell
eps_hat = y - X %*% beta_hat # Residuenvektor vollständiges Modell
eps_hat_0 = y - X_0 %*% beta_hat_0 # Residuenvektor reduziertes Modell
SQT = t(eps_hat_0) %*% eps_hat_0 # total sum of squares
SQW = t(eps_hat) %*% eps_hat # within sum of squares
SQB = SQT - SQW # between sum of squares
DFB = p - 1 # between degrees of freedom
DFW = n - p # within degrees of freedom
MSB = SQB/DFB # mean between sum of squares
MSW = SQW/DFW # mean within sum of squares
Eff = MSB/MSW # F-Teststatistik
pW = 1 - pf(Eff, p-1, n-p) # p-Wert
```

```
# F-Test Evaluation
alpha_0 = 0.05 # Signifikanzlevel
k_alpha_0 = qf(1 - alpha_0, p-1, n-p) # kritischer Wert
if(Eff > k_alpha_0){phi = 1} else {phi = 0} # Testwert

DFB : 2
DFW : 33
SQB : 821.72
SQW : 170.58
MSB : 410.86
MSW : 5.17
F : 79.48
p : 0
phi : 1 2
```

Im vorliegenden Fall würde also die Nullhypothese verworfen werden:

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \in \mathbb{R}^3 \mid \alpha_i = 0 \text{ für } i = 2, 3 \right\} \quad (10.69)$$

Folgender **R**-Code demonstriert die Durchführung und Dokumentation der gleichen Analyse mithilfe der **R**-Funktion `aov()`.

```
D = read.csv("_data/10_Einfaktorielle_Varianzanalyse.csv") # Daten
res.aov = aov(D$dBEDI ~ D$COND, data = D) # Modellformulierung und Modellschätzung
summary(res.aov) # Modellevaluation

Df Sum Sq Mean Sq F value Pr(>F)
D$COND 2 821.7 410.9 79.48 2.41e-13 ***
Residuals 33 170.6 5.2
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 10.6 Literaturhinweise

Die Popularität varianzanalytischer Verfahren wird im Allgemeinen auf Fisher (1925a) und Fisher (1935) zurückgeführt. Everitt und Howell (2005) und Stigler (1986) geben einen kurzen bzw. einen ausführlichen historischen Überblick.

## 10.7 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 und within sum of squares.
14. Geben Sie die Definition des Effektstärkenmaßes  $\eta^2$  an.

15. Wann nimmt das Effektstärkenmaß  $\eta^2$  der EVA seinen Minimalwert an und wie lautet dieser?
16. Wann nimmt das Effektstärkenmaß  $\eta^2$  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.
19. Geben Sie das Theorem zum Zusammenhang von Effektstärkenmaß  $\eta^2$  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.

# 11 Zweifaktorielle Varianzanalyse

## 11.1 Anwendungsszenario

Das klassische Anwendungsszenario einer zweifaktoriellen Varianzanalyse (ZVA) ist ein randomisiertes zweifaktorielles vollständig gekreuztes Studiendesign, bestehend aus einer univariaten abhängigen Variable, bestimmt an randomisierten experimentellen Einheiten, und zwei unabhängigen Variablen, die in der Regel jeweils mindestens zweistufig sind. Im Kontext der Varianzanalyse werden die unabhängigen Variablen bekanntlich auch *Faktoren* und ihre Ausprägungsstufen *Faktorlevel* genannt. Im Falle eines *vollständig gekreuzten Designs* wird jedes Level eines Faktors mit allen Level des jeweils anderen Faktors kombiniert. Die Kombination zweier spezifischer Faktorlevel wird dann auch *Zelle* des Designs genannt. Da den Zellen eines ZVA Designs üblicherweise Gruppen von experimentellen Einheiten zugeordnet werden, werden sie auch oft *Gruppen* oder *experimentelle Bedingungen* genannt.

Spezielle zweifaktorielle Studiendesigns werden üblicherweise anhand ihrer Faktorlevel bezeichnet. Ein  $I \times J$  Studiendesign und seine entsprechende  $I \times J$  zweifaktorielle Varianzanalyse impliziert also, dass  $I$  Faktorlevel des ersten Faktors, im Folgenden als Faktor A bezeichnet, mit  $J$  Faktorleveln des zweiten Faktors, im Folgenden als Faktor B bezeichnet, gekreuzt werden. Folgende Tabelle gibt für einige der möglichen ZVA Designs die Bezeichnung und die Level der jeweiligen Faktoren an.

$2 \times 2$ ZVA	$\Rightarrow$	Faktor A mit Level 1,2	Faktor B mit Level 1,2
$2 \times 3$ ZVA	$\Rightarrow$	Faktor A mit Level 1,2	Faktor B mit Level 1,2,3
$4 \times 2$ ZVA	$\Rightarrow$	Faktor A mit Level 1,2,3,4	Faktor B mit Level 1,2
$3 \times 1$ ZVA	$\Rightarrow$	Faktor A mit Level 1,2,3	Faktor B mit Level 1

Generell sind  $2 \times 2$  ZVA-Designs sehr populär und ermöglichen es, die wesentlichen Charakteristika zweifaktorieller Varianzanalysedesigns zu diskutieren. Wir fokussieren im Folgenden daher weitgehend auf diesen Fall. In Abbildung 11.1 visualisieren wir das konzeptuelle Design einer  $2 \times 2$  ZVA, wobei wir die Zellen des Designs hier und im Folgenden mit A1B1, A1B2, A2B1 und A2B2 bezeichnen.

Abbildung 11.2 visualisiert, dass zu jeder Zelle des Designs eine Gruppen von Datenpunkten korrespondiert. Wir nutzen dabei folgende Indexkonvention:  $y_{ijk}$  bezeichnet den Datenwert der  $k$  ten experimentellen Einheit im  $i$  ten Level von Faktor A und  $j$  ten Level von Faktor B, wobei  $k = 1, \dots, n_{ij}$  und im vorliegenden Fall  $i = 1, 2$  sowie  $j = 1, 2$ . Ist die Anzahl an Datenwerten  $n_{ij}$  in jeder Zelle identisch, so spricht man wiederum von einem balancierten Design.

		<b>Faktor B</b>	
		Level 1	Level 2
<b>Faktor A</b>	Level 1	A1B1	A1B2
	Level 2	A2B1	A2B2

Abbildung 11.1 Konzeptuelles Design einer  $2 \times 2$  ZVA.

		<b>Faktor B</b>	
		Level 1	Level 2
<b>Faktor A</b>	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	$y_{211} = 0.1$ $y_{212} = 0.8$ $y_{213} = 2.7$ $\vdots$ $y_{21n_{21}} = 1.4$	$y_{221} = 7.4$ $y_{222} = 6.2$ $y_{223} = 9.5$ $\vdots$ $y_{22n_{22}} = 6.1$

Abbildung 11.2 Datennotation einer  $2 \times 2$  ZVA.

## Haupteffekte und Interaktionen

Intuitiv ist man in ZVA-Designs an Haupteffekten und an Interaktionen interessiert, die sich zunächst rein deskriptiv auf das Muster der zellspezifischen Gruppenmittelwerte beziehen.

Für den Fall einer  $2 \times 2$  ZVA spricht man dabei intuitiv 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. Weiterhin 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.

Vom Vorliegen einer *Interaktion der Faktoren A und B* schließlich spricht man, 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.

Haupteffekte beziehen sich intuitiv also auf (marginale) Unterschiede (Differenzen), während sich Interaktionen auf Unterschiede von Unterschieden (Differenzen von Differenzen) beziehen. Dabei besagt das Vorliegen einer Interaktion also lediglich, dass sich die Unterschiede der Gruppenmittelwerte zwischen den Leveln eines experimentellen Faktors in Abhängigkeit von den Leveln des anderen experimentellen Faktors ändern, impliziert aber keine Aussage darüber, warum dies so ist. In anderen Worten sind Haupteffekte und Interaktionen bei Varianzanalysen Datenmuster, aber keine wissenschaftlichen Theorien. Die frequentenistisch-inferenzstatistische Absicherung solcher Datenmuster ist das Thema dieses Kapitels.

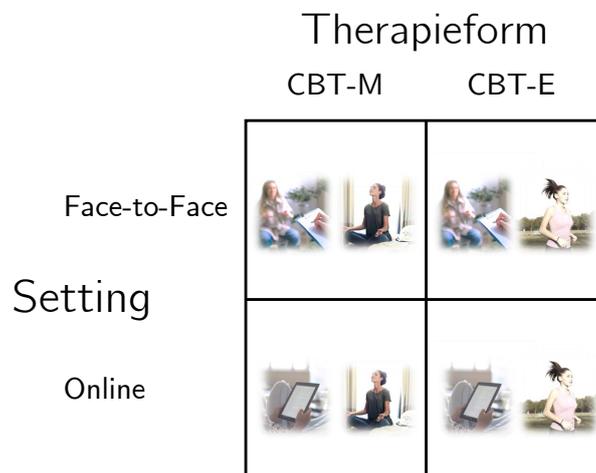


Abbildung 11.3 Konzeptuelles Design des Anwendungsbeispiels

## 11.2 Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel aus dem Bereich der Interventionsevaluation betrachten wir eine BDI-Differenzwertanalyse für je zwei Settings (Face-to-Face und Online)

und Varianten (Mindfulness und Exercise) der kognitiven Verhaltenstherapie (CBT, siehe Abbildung 11.3).

Setting sei Faktor A mit Leveln (1) Face-to-face (F2F) und (2) Online (ONL) und Variante sei Faktor B mit Leveln (1) Mindfulness (MND) und (2) Exercise (EXC). Jeder der durch Kreuzung der Faktorlevel entstehenden Interventionsbedingungen sei eine Gruppe (Stichprobe) von  $n_{ij} := 12$  Patient:innen zugeordnet und für jede Patient:in sei die Differenz der BDI-Werte zwischen vor und nach der Intervention bestimmt worden und mit dBDI bezeichnet.

**Tabelle 11.1** Pre-Post-BDI-Differenzwerte der Faktorlevelkombinationen

Setting	Variant	dBDI
F2F	MND	12
F2F	MND	11
F2F	MND	8
F2F	MND	8
F2F	MND	12
F2F	MND	12
F2F	MND	9
F2F	MND	10
F2F	MND	12
F2F	MND	13
F2F	MND	10
F2F	MND	9
F2F	EXC	14
F2F	EXC	11
F2F	EXC	15
F2F	EXC	16
F2F	EXC	15
F2F	EXC	19
F2F	EXC	16
F2F	EXC	14
F2F	EXC	11
F2F	EXC	15
F2F	EXC	15
F2F	EXC	17
ONL	MND	4
ONL	MND	10
ONL	MND	12
ONL	MND	13
ONL	MND	12
ONL	MND	12
ONL	MND	13
ONL	MND	10
ONL	MND	10
ONL	MND	13
ONL	MND	4
ONL	MND	8
ONL	EXC	16
ONL	EXC	19
ONL	EXC	14
ONL	EXC	17
ONL	EXC	17
ONL	EXC	16
ONL	EXC	13
ONL	EXC	16
ONL	EXC	20
ONL	EXC	13
ONL	EXC	16
ONL	EXC	13

Tabelle 11.1 zeigt einen Beispieldatensatz. Dabei repräsentiert jede Zeile eine Patient:in, die Spalte **Setting** bezeichnet das dieser Patient:in entsprechende Level des Therapiesettingfaktors und die Spalte **Variant** das dieser Patient:in entsprechende Level des Therapievariantenfaktors. Der entsprechende dBDI-Zeilenswert bildet den PBDI-Differenzwert der entsprechenden Patient:in ab. Hohe Werte zeigen dabei eine starke Verbesserung, geringe Werte eher eine leichte Verbesserung der Depressionssymptomatik an.

Wir visualisieren an dieser Stelle zwei häufig genutzte Darstellungsformen für Deskriptivstatistiken (hier Gruppenmittelwerte und Standardabweichungen) in  $2 \times 2$  ZVA-Szenarien. Abbildung 11.4 zeigt die Faktorlevel-Kombinations-spezifischen Gruppenmittelwerte und ihre Standardabweichungen als Balkendiagramm. Abbildung 11.5 zeigt die gleichen Daten als Liniendiagramm mit Fehlerbalken. Dabei entsprechen die zwei Linien den beiden

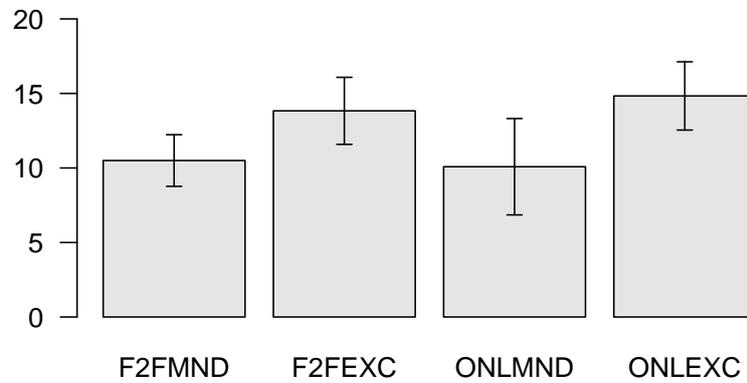


Abbildung 11.4 Visualisierung des Beispieldatensatzes als Balkendiagramm

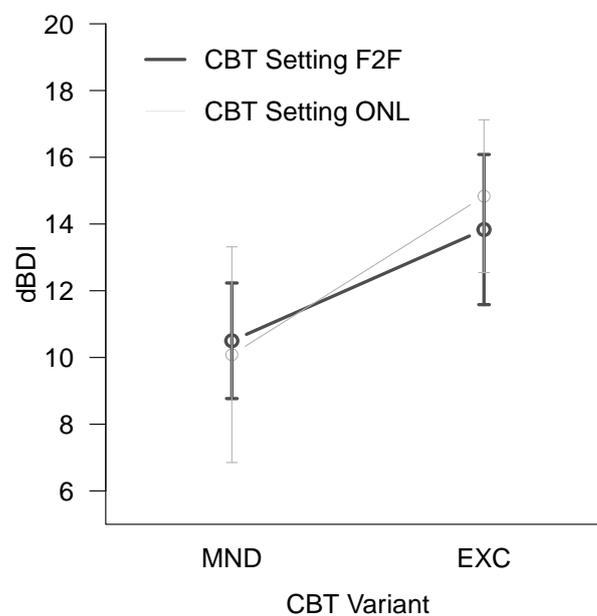


Abbildung 11.5 Visualisierung des Beispieldatensatzes als Liniendiagramm

Leveln des Settingfaktors F2F (dunkelgraue Linie) und ONL (hellgraue Linie). Die auf der x-Achse abgetragenen Werte dagegen entsprechen den beiden Leveln des Variantenfaktors MND und EXC. Für das Verständnis dieser Art der Visualisierung ist es sicherlich hilfreich, die Korrespondenz zwischen den Deskriptivstatistiken in Abbildung 11.4 und Abbildung 11.5 einmal explizit durch nachzeichnen herzustellen.

## 11.3 Modellformulierung

Um nun den Spezialfall des ALMs für das ZVA-Szenario zu diskutieren, gehen wir in zwei Schritten vor. In einem ersten Schritt führen wir mit dem *Modell der additiven ZVA* ein Modell ein, das lediglich die Haupteffekte der Faktoren, nicht aber ihre Interaktion modellieren kann. Entsprechend können auf Grundlage dieses Modells Haupteffektparameter geschätzt und inferenzstatistisch evaluiert werden.

In einem zweiten Schritt führen wir mit dem Modell der ZVA mit Interaktion ein Modell ein, das sowohl die Haupteffekte der Faktoren als auch ihre Interaktion modelliert und auf dessen Grundlage entsprechend sowohl Haupteffekts- als auch Interaktionsparameter geschätzt und inferenzstatistisch evaluiert werden können. In beiden Fällen ist das zentrale Thema der Modellformulierung wie bei der EVA von einer Intuition zur Modellierung der Erwartungswertparameter der Zellen des ZVA-Szenarios zu einer nicht überparameterisierten Darstellung mithilfe von Effektparametern zu kommen.

### 11.3.1 Modell der additiven ZVA

In Analogie zur EVA möchte man im Modell der additiven ZVA die Gruppenerwartungswerte  $\mu_{ij}$  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 dabei den gruppenunspezifischen Erwartungswertparameter mit  $\mu_0$ , den Effekt von Level  $i$  von Faktor A mit  $\alpha_i$  und den Effekt von Level  $j$  von Faktor B mit  $\beta_j$  ( $\beta_j$  bezeichnet hier also nicht den  $j$  ten Eintrag des Betaparametervektors). Dann ergibt sich zum Beispiel für  $I := J := 2$

$$\begin{array}{l|l} \mu_{11} := \mu_0 + \alpha_1 + \beta_1 & \mu_{12} := \mu_0 + \alpha_1 + \beta_2 \\ \mu_{21} := \mu_0 + \alpha_2 + \beta_1 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 \end{array} \quad (11.1)$$

Wie im Falle der EVA ist diese Darstellung der Gruppenerwartungswerte  $\mu_{ij}$  allerdings überparameterisiert, speziell werden die vier Erwartungswertparameter  $\mu_{11}$ ,  $\mu_{12}$ ,  $\mu_{21}$  und  $\mu_{22}$  in obiger Formulierung durch die fünf Parameter  $\mu_0$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$  und  $\beta_2$  dargestellt. Um eine eindeutige Darstellung der  $\mu_{ij}$  zu gewährleisten, bietet sich auch hier die Restriktion an, den Effekt des ersten Levels jedes Faktors als Null zu definieren

$$\alpha_1 := \beta_1 := 0 \quad (11.2)$$

und damit die Faktorlevelkombination A1B1 als Referenzgruppe zu etablieren. Es ergibt sich somit zum Beispiel für  $I := J := 2$ :

$$\begin{array}{l|l} \mu_{11} := \mu_0 & \mu_{12} := \mu_0 + \beta_2 \\ \mu_{21} := \mu_0 + \alpha_2 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 \end{array} \quad (11.3)$$

Hier werden nun also die vier Erwartungswertparameter  $\mu_{11}$ ,  $\mu_{12}$ ,  $\mu_{21}$  und  $\mu_{22}$  durch nur drei Effektparameter  $\mu_0$ ,  $\alpha_2$  und  $\beta_2$  dargestellt. (In diesem Sinne ist das Modell der additiven ZVA sogar unterparameterisiert, was im folgenden Abschnitt die Einführung eines Interaktionsparameters erlaubt.) Allerdings ändern sich bei dieser *Effektdarstellung des Modells der additiven  $2 \times 2$  ZVA mit Referenzgruppe* analog zur EVA die Interpretationen der Parameter  $\mu_0$ ,  $\alpha_2$ ,  $\beta_2$  im Vergleich zum überparameterisierten Fall ohne Referenzgruppe:  $\mu_0$  entspricht dem Erwartungswert der Faktorlevel-Kombination A1B1,  $\alpha_2$  der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A und  $\beta_2$  der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B. Wir fassen obige Überlegungen in folgender Definition zusammen.

**Definition 11.1** (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 innerhalb des  $i$ ten Levels von Faktor A und des  $j$ ten Levels von Faktor B in einem ZVA-Anwendungsszenario modelliert. Dann hat das *Modell der additiven ZVA mit Referenzgruppe* die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \text{ mit } \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.4)$$

und die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.5)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j \text{ für } i = 1, \dots, I, j = 1, \dots, J \text{ mit } \alpha_1 := \beta_1 := 0 \quad (11.6)$$

und  $\sigma^2 > 0$ .

•

Wir verzichten in Definition 11.1 auf die Angabe einer allgemeinen Designmatrixform, die wir im Folgenden lediglich für den  $2 \times 2$  ZVA Spezialfall betrachten.

Die Expressivität des Modells der additiven ZVA ist, wie schon oben betont, auf das Abbilden von Haupteffekten beschränkt. Wir verdeutlichen dies in folgenden Parameterbeispielen und der zugehörigen Abbildung 11.6.

### Beispiel (1)

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 1$  und  $\beta_2 := 0$ , der Effektparameter für den Haupteffekt von Faktor A sei also von Null verschieden, der Effektparameter für den Haupteffekt von Faktor B dagegen gleich Null. Dann ergibt sich für die Gruppenerwartungswerte

$$\begin{array}{l|l} \mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = 1 & \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 0 = 1 \\ \mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 1 + 1 + 0 = 2 & \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 1 + 1 + 0 = 2 \end{array} \quad (11.7)$$

Abbildung 11.6 A zeigt Visualisierungen dieses Erwartungswertmusters in der Balkendiagramm- und Liniendiagrammform.

### Beispiel (2)

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 0$  und  $\beta_2 := 1$ , der Effektparameter für den Haupteffekt von Faktor A sei also gleich Null, der Effektparameter für den Haupteffekt von Faktor B dagegen von Null verschieden. Dann ergibt sich für die Gruppenerwartungswerte

$$\frac{\mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = 1 \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 1 = 2}{\mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 1 + 0 + 0 = 1 \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 1 + 0 + 1 = 2} \quad (11.8)$$

Abbildung 11.6 B zeigt Visualisierungen dieses Erwartungswertmusters in der Balkendiagramm- und Liniendiagrammform.

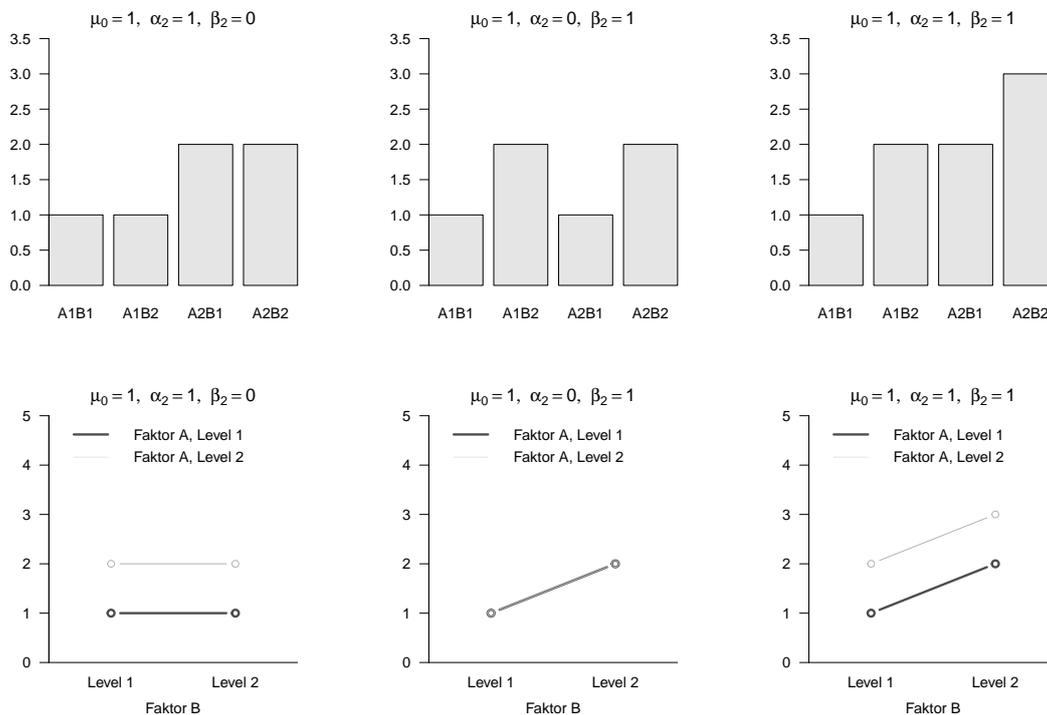
**Beispiel (3)**

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 1$  und  $\beta_2 := 1$ , die Effektparameter für die Haupteffekte beider Faktoren seien also von Null verschieden. Dann ergibt sich für die Gruppenerwartungswerte

$$\frac{\mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = 1 \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 1 = 2}{\mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 1 + 1 + 0 = 2 \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 1 + 1 + 1 = 3} \quad (11.9)$$

Abbildung 11.6 C zeigt Visualisierungen dieses Erwartungswertmusters in der Balkendiagramm- und Liniendiagrammform.

In Hinblick auf die Liniendiagrammform der Gruppenerwartungswerte der additiven ZVA fällt auf, dass die den Leveln von Faktor A entsprechenden Linien immer parallel sind.



**Abbildung 11.6** Gruppenerwartungswerte der Beispielparameterszenarien in Balken- und Liniendiagrammform im Modell der additiven ZVA

Für den Fall der  $2 \times 2$  ZVA gibt Definition 11.2 abschließend das Modell der additiven ZVA mit Referenzgruppe inklusive seiner Designmatrixform an.

**Definition 11.2** (Modell der additiven  $2 \times 2$  ZVA mit Referenzgruppe).  $y_{ijk}$  mit  $i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}$  sei die Zufallsvariable, die den  $k$ ten Datenpunkt innerhalb des  $i$ ten Levels von Faktor A und des  $j$ ten Levels von Faktor B in einem  $2 \times 2$  ZVA-Anwendungsszenario modelliert. Dann hat mit  $\sigma^2 > 0$  das *Modell der additiven  $2 \times 2$  ZVA mit Referenzgruppe* die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \text{ mit } \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. f\u00fcr } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}, \quad (11.10)$$

die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. f\u00fcr } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij} \quad (11.11)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j \text{ f\u00fcr } i = 1, 2, j = 1, 2 \text{ mit } \alpha_1 := \beta_1 := 0, \quad (11.12)$$

sowie die Designmatrixform

$$y \sim N(X\beta, \sigma^2 I_n) \quad (11.13)$$

mit

$$y := \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times 3}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (11.14)$$

•

### 11.3.2 Modell der ZVA mit Interaktion

In der ZVA mit Interaktion m\u00f6chte man die Gruppenerwartungswerte  $\mu_{ij}$  mit  $i = 1, \dots, I$  f\u00fcr die Level von Faktor A und  $j = 1, \dots, J$  f\u00fcr 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 dazu den gruppenunspezifischen Erwartungswertparameter mit  $\mu_0$ , den Effekt von Level  $i$  von Faktor A mit  $\alpha_i$ , den Effekt von Level  $j$  von Faktor B mit  $\beta_j$ , und die Interaktion von Level  $i$  von Faktor A mit Level  $j$  von Faktor B mit  $\gamma_{ij}$ . Dann ergibt sich zum Beispiel f\u00fcr  $I := J := 2$

$$\begin{array}{l|l} \mu_{11} := \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} & \mu_{12} := \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} \\ \mu_{21} := \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} \end{array} \quad (11.15)$$

Wie in der die additiven ZVA ist diese Darstellung der Gruppenerwartungswerte  $\mu_{ij}$  multipel \u00fcberparameterisiert, speziell werden die vier Erwartungswertparameter  $\mu_{11}, \mu_{12}, \mu_{21}$

und  $\mu_{22}$  in obiger Formulierung durch die neun Parameter  $\mu_0, \alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_{11}, \gamma_{12}, \gamma_{21}$  und  $\gamma_{22}$  dargestellt. Um eine eindeutige Darstellung der  $\mu_{ij}$  zu gewährleisten, bietet sich auch hier die Restriktion an, den Effekt des ersten Levels jedes Faktors und jeder Interaktion mit dem ersten Level eines Faktors als Null zu definieren

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \dots, I, j = 1, \dots, J \quad (11.16)$$

und damit die Faktorlevel-Kombination A1B1 als Referenzgruppe zu etablieren. Es ergibt sich somit zum Beispiel für  $I := J := 2$ :

$$\begin{array}{c|c} \mu_{11} := \mu_0 & \mu_{12} := \mu_0 + \beta_2 \\ \hline \mu_{21} := \mu_0 + \alpha_2 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} \end{array} \quad (11.17)$$

Hier werden nun also die vier Erwartungswertparameter  $\mu_{11}, \mu_{12}, \mu_{21}$  und  $\mu_{22}$  durch die vier Effektparameter  $\mu_0, \alpha_2, \beta_2$  und  $\gamma_{22}$  dargestellt. Natürlich ändern sich auch bei dieser Effektdarstellung des Modells der  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe die Interpretationen der Parameter  $\mu_0, \alpha_2, \beta_2, \gamma_{22}$  im Vergleich zum überparameterisierten Fall ohne Referenzgruppe:  $\mu_0$  entspricht dem Erwartungswert der Faktorlevel-Kombination A1B1,  $\alpha_2$  der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A,  $\beta_2$  der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B und  $\gamma_{22}$  der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B im Unterschied zum Übergang von Level 1 zu Level 2 von Faktor A. Wir fassen obige Überlegungen in folgender Definition zusammen, wobei wir wiederum auf die Angabe einer allgemeinen Designmatrixform verzichten wollen.

**Definition 11.3** (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

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \text{ mit } \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.18)$$

und die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.19)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j + \gamma_{ij} \quad (11.20)$$

sowie

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \dots, I, j = 1, \dots, J \quad (11.21)$$

und  $\sigma^2 > 0$ .

•

Die Expressivität des Modells der ZVA mit Interaktion entspricht nun dem Betrachten von Haupteffekten und Interaktion wie einleitend zum Anwendungsszenario der ZVA diskutiert. Wir wollen dies mithilfe folgender Parameterbeispiele und der zugehörigen Abbildung 11.7 verdeutlichen.

### Beispiel (1)

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 0$ ,  $\beta_2 := 0$  und  $\gamma_{22} = 2$ , die Effektparameter für die Haupteffekte beider Faktoren seien also gleich Null, aber der Effektparameter für die Interaktion der Faktoren positiv. Dann ergibt sich für die Gruppenerwartungswerte

$$\frac{\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1}{\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 0 + 0 = 1} \mid \frac{\mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 0 + 0 = 1}{\mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 0 + 0 + 2 = 3} \quad (11.22)$$

Abbildung 11.7 A zeigt Visualisierungen dieses Erwartungswertmusters in Balkendiagramm- und Liniendiagrammform.

### Beispiel (2)

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 1$ ,  $\beta_2 := 1$  und  $\gamma_{22} = -2$ . Die Effektparameter für die Haupteffekte beider Faktoren seien also gleich und der Effektparameter für die Interaktion der Faktoren negativ. Dann ergibt sich für die Gruppenerwartungswerte

$$\frac{\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1}{\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 1 + 0 = 2} \mid \frac{\mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 1 + 0 = 2}{\mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 1 + 1 - 2 = 1} \quad (11.23)$$

Abbildung 11.7 B zeigt Visualisierungen dieses Erwartungswertmusters in Balkendiagramm- und Liniendiagrammform.

### Beispiel (3)

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 1$ ,  $\beta_2 := 0$  und  $\gamma_{22} = 1$ , der Effektparameter für Faktor A sei also positiv, der Effektparameter für Faktor B gleich Null und der Effektparameter für die Interaktion der Faktoren positiv. Dann ergibt sich für die Gruppenerwartungswerte

$$\frac{\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1}{\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 1 + 0 + 0 = 2} \mid \frac{\mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 0 + 0 = 1}{\mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 1 + 0 + 1 = 3} \quad (11.24)$$

Abbildung 11.7 C zeigt Visualisierungen dieses Erwartungswertmusters in Balkendiagramm- und Liniendiagrammform.

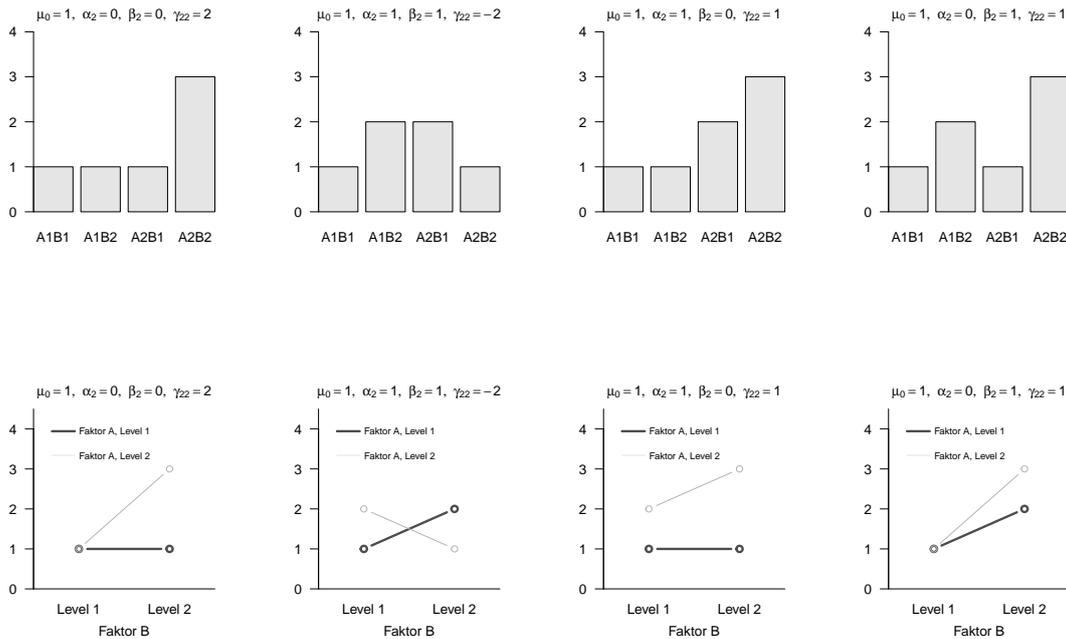
### Beispiel (4)

Es seien  $\mu_0 := 1$ ,  $\alpha_2 := 0$ ,  $\beta_2 := 1$  und  $\gamma_{22} = 1$ , der Effektparameter für Faktor A sei also gleich Null, der Effektparameter für Faktor B positiv und der Effektparameter für die Interaktion der Faktoren wiederum positiv. Dann ergibt sich für die Gruppenerwartungswerte

$$\frac{\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1}{\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 0 + 0 = 1} \mid \frac{\mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 1 + 0 = 2}{\mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 0 + 1 + 1 = 3} \quad (11.25)$$

Abbildung 11.7 D zeigt Visualisierungen dieses Erwartungswertmusters in Balkendiagramm- und Liniendiagrammform.

Die obigen Beispiele zeigen, dass die Expressivität der ZVA mit Interaktion im Sinne der durch das Modell abzubildenden Datenmuster ungleich höher als die Expressivität des rein additiven Modells der ZVA ist. Insbesondere bilden obige vier Beispiele bei Weitem nicht den gesamten kombinatorischen qualitativen Parameterraum der ZVA mit Interaktion ab. Hinsichtlich der Liniendiagrammvisualisation der Gruppenerwartungswertparameter bei ZVA mit Interaktion fällt auf, dass ein von Null verschiedener Interaktionsterm immer impliziert, dass die die Level von Faktor A repräsentierenden Linien nicht mehr parallel verlaufen.



**Abbildung 11.7** Gruppenerwartungswerte der Beispielparameterszenarien in Balken- und Liniendiagrammform im Modell der ZVA mit Interaktion

Abschließend geben wir in Definition 11.4 den Spezialfall des Modells der  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe inklusiver seiner Designmatrixform an.

**Definition 11.4** (Modell der  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe).  $y_{ijk}$  mit  $i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}$  sei die Zufallsvariable, die den  $k$ ten Datenpunkt innerhalb des  $i$ ten Levels von Faktor A und des  $j$ ten Levels von Faktor B in einem  $2 \times 2$  ZVA Anwendungsszenario modelliert. Dann hat das Modell der  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe für  $\sigma^2 > 0$  die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \text{ mit } \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}, \quad (11.26)$$

die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij} \quad (11.27)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j + \gamma_{ij} \quad (11.28)$$

sowie

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \dots, I, j = 1, \dots, J \quad (11.29)$$

und für  $n := \sum_{i=1}^2 \sum_{j=1}^2 n_{ij}$  die Designmatrixform

$$y \sim N(X\beta, \sigma^2 I_n) \quad (11.30)$$

mit

$$y := \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times 4}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (11.31)$$

•

### Beispiel

Als konkretes Beispiel für die Designmatrixform des in Definition 11.4 definierten Modells betrachten wir das Szenario  $n_{ij} := 4$ , also  $n = 16$ . Dann gilt

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{16}, \sigma^2 I_{16}) \quad (11.32)$$

mit

$$y := \begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{124} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{224} \end{pmatrix}, X := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{16 \times 4}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (11.33)$$

Folgender **R**-Code erlaubt die Realisierung von Daten im Falle dieses Beispiels.

```

library(MASS) # multivariate Normalverteilung
I = 2 # Anzahl Level Faktor A
J = 2 # Anzahl Level Faktor B
n_ij = 4 # Anzahl von Datenpunkten der (i,j)-ten Gruppe
n = I*J*n_ij # Anzahl Datenpunkte
p = 1 + (I-1)+(J-1)+(I*J-3) # Anzahl Parameter
D = matrix(c(1,0,0,0, # prototypische Designmatrix für balancierte Designs
            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) # prototypischer Zellenvektor für balancierte Designs
X = kronecker(D,C) # Kroneckerprodukt-Designmatrix-Erzeugung
I_n = diag(n) # n x n Einheitsmatrix
beta = matrix(c(1,1,1,1), nrow = p) # \beta = (\mu_0, \alpha_2, \alpha_3, \alpha_4)
sigsqr = 10 # \sigma^2
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y

```

## 11.4 Modellschätzung

Wir betrachten im Folgenden die Betaparameterschätzer im additiven  $2 \times 2$  ZVA Modell mit Referenzgruppe und im  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe. Dabei stellt sich wenig überraschend heraus, dass sich die Schätzer für die Komponenten des Betaparameters der in Definition 11.2 und Definition 11.4 formulierten Modelle als gewichtete Summen der gruppenspezifischen Stichprobenmittel ergeben. Die genaue Form dieser gewichteten Summen ist Inhalt der beiden folgenden Theoreme. Der Einfachheit halber gehen wir dabei jeweils von einem balancierten Design aus. Wir beginnen mit der Betaparameterschätzung im additiven  $2 \times 2$  ZVA Modell mit Referenzgruppe.

**Theorem 11.1** (Betaparameterschätzung im additiven  $2 \times 2$  ZVA Modell mit Referenzgruppe). *Gegeben sei die Designmatrixform eines balancierten additiven  $2 \times 2$  ZVA-Modells mit Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer*

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4}\bar{y}_{11} + \frac{1}{4}(\bar{y}_{12} + \bar{y}_{21}) - \frac{1}{4}\bar{y}_{22} \\ \frac{1}{2}(\bar{y}_{21} + \bar{y}_{22}) - \frac{1}{2}(\bar{y}_{11} + \bar{y}_{12}) \\ \frac{1}{2}(\bar{y}_{12} + \bar{y}_{22}) - \frac{1}{2}(\bar{y}_{11} + \bar{y}_{21}) \end{pmatrix}, \quad (11.34)$$

wobei

$$\bar{y}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \text{ für } 1 \leq i, j \leq 2 \quad (11.35)$$

das Stichprobenmittel der  $(i, j)$ -ten Gruppe des  $2 \times 2$  ZVA-Designs bezeichnet.

◦

*Beweis.* Wir bestimmen zunächst  $X^T y$ ,  $X^T X$  und  $(X^T X)^{-1}$  bei konstantem  $n_{ij}$  für  $1 \leq i, j \leq 2$ .

$$\begin{aligned}
 X^T y &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \end{pmatrix}
 \end{aligned} \tag{11.36}$$

$$\begin{aligned}
 X^T X &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix} \\
 &= n_{ij} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}
 \end{aligned} \tag{11.37}$$

Ohne Beweis halten wir weiterhin fest, dass

$$(X^T X)^{-1} = n_{ij} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}^{-1} = \frac{1}{n_{ij}} \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix}. \tag{11.38}$$

Es ergibt sich also

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \frac{1}{n_{ij}} \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \end{pmatrix}. \tag{11.39}$$

Damit ergibt sich dann

$$\begin{aligned}
 \hat{\mu}_0 &= \frac{1}{n_{ij}} \left( \frac{3}{4} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} - \frac{1}{2} \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} - \frac{1}{2} \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \right) \\
 &= \frac{1}{n_{ij}} \left( \frac{3}{4} \sum_{k=1}^{n_{11}} y_{11k} + \frac{3}{4} \sum_{k=1}^{n_{12}} y_{12k} + \frac{3}{4} \sum_{k=1}^{n_{21}} y_{21k} + \frac{3}{4} \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &\quad + \frac{1}{n_{ij}} \left( -\frac{1}{2} \sum_{k=1}^{n_{21}} y_{21k} - \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} - \frac{1}{2} \sum_{k=1}^{n_{12}} y_{12k} - \frac{1}{2} \sum_{k=1}^{n_{11}} y_{11k} \right) \\
 &= \frac{1}{n_{ij}} \left( \frac{3}{4} \sum_{k=1}^{n_{11}} y_{11k} + \frac{1}{4} \sum_{k=1}^{n_{12}} y_{12k} + \frac{1}{4} \sum_{k=1}^{n_{21}} y_{21k} - \frac{1}{4} \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &= \frac{3}{4} \bar{y}_{11} + \frac{1}{4} (\bar{y}_{12} + \bar{y}_{21}) - \frac{1}{4} \bar{y}_{22}
 \end{aligned} \tag{11.40}$$

sowie

$$\begin{aligned}
 \hat{\alpha}_2 &= \frac{1}{n_{ij}} \left( -\frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} + \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \right) \\
 &= \frac{1}{n_{ij}} \left( -\frac{1}{2} \sum_{k=1}^{n_{11}} y_{11k} - \frac{1}{2} \sum_{k=1}^{n_{12}} y_{12k} - \frac{1}{2} \sum_{k=1}^{n_{21}} y_{21k} - \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} + \sum_{k=1}^{n_{21}} y_{21k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left( -\frac{1}{2} \sum_{k=1}^{n_{11}} y_{11k} - \frac{1}{2} \sum_{k=1}^{n_{12}} y_{12k} + \frac{1}{2} \sum_{k=1}^{n_{21}} y_{21k} + \frac{1}{2} \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &= \frac{1}{2} (\bar{y}_{21} + \bar{y}_{22}) - \frac{1}{2} (\bar{y}_{11} + \bar{y}_{12})
 \end{aligned} \tag{11.41}$$

und analog für  $\hat{\beta}_2$ . □

## Beispiel

Wir verdeutlichen den in Theorem 11.1 formulierten Zusammenhang zwischen den Beta-parameterschätzerkomponenten und den gruppenspezifischen Stichprobenmitteln mithilfe folgenden **R**-Codes am Beispieldatensatz.

```

# Datenreformatierung
D = read.csv("_data/11_Zweifaktorielle_Varianzanalyse.csv")
A1B1 = D$dBDI[D$Setting == "F2F" & D$Variant == "MND" ]
A1B2 = D$dBDI[D$Setting == "F2F" & D$Variant == "EXC" ]
A2B1 = D$dBDI[D$Setting == "ONL" & D$Variant == "MND" ]
A2B2 = D$dBDI[D$Setting == "ONL" & D$Variant == "EXC" ]

# Datensatz
# face-to-face, mindfulness
# face-to-face, exercise
# online, mindfulness
# online, exercise

# Datenmatrix für Gruppenmittelwerte
n_ij = length(A1B1)
Y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n_ij)
bar_y = colMeans(Y)

# Anzahl Datenpunkte pro Gruppe
# Datenmatrix
# Zellenmittelwerte

# 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,
             1,0,1,
             1,1,0,
             1,1,1),
           nrow = p,
           byrow = TRUE)

# Anzahl Level Faktor A (Therapiesetting)
# Anzahl Level Faktor B (Therapievariante)
# Anzahl Datenpunkte
# Anzahl Parameter
# prototypische Designmatrix für balancierte Designs

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)

# prototypischer Zellenvektor für balancierte Designs
# Kroneckerprodukt-Designmatrix-Erzeugung
# Datenvektor
# Betaparameterschätzer
# Residuenvektor
# Varianzparameterschätzer

hat{beta} : 10.15 0.29 5.04
hat{sigsqr} : 6.07
bar{y}_11,bar{y}_12,bar{y}_21,bar{y}_22 : 10.5 14.83 10.08 15.83
3/4bar{y}_11 + 1/4(bar{y}_12 + bar{y}_21) - 1/4bar{y}_22 : 10.15
1/2(bar{y}_21 + bar{y}_22) - 1/2(bar{y}_11 + bar{y}_12) : 0.29
1/2(bar{y}_12 + bar{y}_22) - 1/2(bar{y}_11 + bar{y}_21) : 5.04

```

Für die Betaparameterschätzung im  $2 \times 2$  ZVA Modell mit Interaktion und Referenzgruppe ergibt sich folgendes Theorem.

**Theorem 11.2** (Betaparameterschätzung im  $2 \times 2$  ZVA Modell mit Interaktion und Referenzgruppe). *Gegeben sei die Designmatrixform eines balancierten  $2 \times 2$  ZVA-Modells mit Interaktion und Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer*

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \begin{pmatrix} \bar{y}_{11} \\ \bar{y}_{21} - \bar{y}_{11} \\ \bar{y}_{12} - \bar{y}_{11} \\ \bar{y}_{11} + \bar{y}_{22} - \bar{y}_{12} - \bar{y}_{21} \end{pmatrix}, \quad (11.42)$$

wobei

$$\bar{y}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \text{ für } 1 \leq i, j \leq 2 \quad (11.43)$$

das Stichprobenmittel der  $(i, j)$ ten Gruppe des  $2 \times 2$  ZVA-Designs bezeichnet.

◦

*Beweis.* Wir bestimmen zunächst  $X^T y$ ,  $X^T X$  und  $(X^T X)^{-1}$  bei konstantem  $n_{ij}$  für  $1 \leq i, j \leq 2$ .

$$X^T y = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ y_{121} \\ \vdots \\ y_{12n_{12}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ y_{221} \\ \vdots \\ y_{22n_{22}} \end{pmatrix} \quad (11.44)$$

$$= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \\ \sum_{k=1}^{n_{22}} y_{22k} \end{pmatrix}$$

$$\begin{aligned}
X^T X &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \\
&= n_{ij} \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}
\end{aligned} \tag{11.45}$$

Ohne Beweis halten wir weiterhin fest, dass

$$(X^T X)^{-1} = n_{ij} \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}^{-1} = \frac{1}{n_{ij}} \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 4 \end{pmatrix}. \tag{11.46}$$

Es ergibt sich also

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \frac{1}{n_{ij}} \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 4 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} \\ \sum_{k=1}^{n_{22}} y_{22k} \end{pmatrix}. \tag{11.47}$$

Damit ergibt sich dann

$$\begin{aligned}
\hat{\mu}_0 &= \frac{1}{n_{ij}} \left( \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} - \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} - \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\
&= \frac{1}{n_{ij}} \left( \sum_{k=1}^{n_{11}} y_{11k} + \sum_{k=1}^{n_{12}} y_{12k} + \sum_{k=1}^{n_{21}} y_{21k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\
&+ \frac{1}{n_{ij}} \left( - \sum_{k=1}^{n_{21}} y_{21k} - \sum_{k=1}^{n_{22}} y_{22k} - \sum_{k=1}^{n_{12}} y_{12k} - \sum_{k=1}^{n_{22}} y_{22k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\
&= \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} y_{11k} \\
&= \bar{y}_{11}
\end{aligned} \tag{11.48}$$

sowie

$$\begin{aligned}
\hat{\alpha}_2 &= \frac{1}{n_{ij}} \left( - \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} + 2 \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} + 1 \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} - 2 \sum_{k=1}^{n_{22}} y_{22k} \right) \\
&= \frac{1}{n_{ij}} \left( - \sum_{k=1}^{n_{11}} y_{11k} - \sum_{k=1}^{n_{12}} y_{12k} - \sum_{k=1}^{n_{21}} y_{21k} - \sum_{k=1}^{n_{22}} y_{22k} \right) \\
&+ \frac{1}{n_{ij}} \left( 2 \sum_{k=1}^{n_{21}} y_{21k} + 2 \sum_{k=1}^{n_{22}} y_{22k} + \sum_{k=1}^{n_{12}} y_{12k} + \sum_{k=1}^{n_{22}} y_{22k} - 2 \sum_{k=1}^{n_{22}} y_{22k} \right) \\
&= \frac{1}{n_{ij}} \left( \sum_{k=1}^{n_{21}} y_{21k} - \sum_{k=1}^{n_{11}} y_{11k} \right) \\
&= \bar{y}_{21} - \bar{y}_{11}
\end{aligned} \tag{11.49}$$

und analog für  $\hat{\beta}_2$ . Schließlich ergibt sich

$$\begin{aligned}
 \hat{\gamma}_{22} &= \frac{1}{n_{ij}} \left( \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} y_{ijk} - 2 \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} y_{2jk} - 2 \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} y_{i2k} + 4 \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left( \sum_{k=1}^{n_{11}} y_{11k} + \sum_{k=1}^{n_{12}} y_{12k} + \sum_{k=1}^{n_{21}} y_{21k} + \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &+ \frac{1}{n_{ij}} \left( -2 \sum_{k=1}^{n_{21}} y_{21k} - 2 \sum_{k=1}^{n_{22}} y_{22k} - 2 \sum_{k=1}^{n_{12}} y_{12k} - 2 \sum_{k=1}^{n_{22}} y_{22k} + 4 \sum_{k=1}^{n_{22}} y_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left( \sum_{k=1}^{n_{11}} y_{11k} + \sum_{k=1}^{n_{22}} y_{22k} - \sum_{k=1}^{n_{12}} y_{12k} - \sum_{k=1}^{n_{21}} y_{21k} \right) \\
 &= \bar{y}_{11} + \bar{y}_{22} - \bar{y}_{12} - \bar{y}_{21}.
 \end{aligned}
 \tag{11.50}$$

□

Auch hier verdeutlichen wir den in obigem Theorem formulierten Zusammenhang zwischen den Betaparameterschätzerkomponenten und den gruppenspezifischen Stichprobenmitteln anhand des Beispieldatensatzes mithilfe des folgenden **R**-Codes.

```

# Datenreformatierung
D = read.csv("data/11_Zweifaktorielle_Varianzanalyse.csv") # Datensatz
A1B1 = D$BDI[D$Setting == "F2F" & D$Variant == "MND" ] # face-to-face, mindfulness
A1B2 = D$BDI[D$Setting == "F2F" & D$Variant == "EXC" ] # face-to-face, exercise
A2B1 = D$BDI[D$Setting == "ONL" & D$Variant == "MND" ] # online, mindfulness
A2B2 = D$BDI[D$Setting == "ONL" & D$Variant == "EXC" ] # online, exercise

# 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 (Therapiesetting)
J = 2 # Anzahl Level Faktor B (Therapievariante)
n = I*J*n_ij # Anzahl Datenpunkte
p = 1 + (I-1)+(J-1)+(I*J-3) # Anzahl Parameter
D = matrix(c(1,0,0,0,
             1,0,1,0,
             1,1,0,0,
             1,1,1,1),
           nrow = p,
           byrow = TRUE) # prototypische Designmatrix für balancierte Designs

C = matrix(rep(1,n_ij), nrow = n_ij) # prototypischer Zellenvektor für balancierte Designs
X = kronecker(D,C) # Kroneckerprodukt-Designmatrix-Erzeugung
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

hat{beta} : 10.5 -0.42 4.33 1.42
hat{sigsqr} : 5.94
bar{y}_11,bar{y}_12,bar{y}_21,bar{y}_22 : 10.5 14.83 10.08 15.83
bar{y}_11 : 10.5
bar{y}_21 - bar{y}_11 : -0.42
bar{y}_12 - bar{y}_11 : 4.33
bar{y}_11 + bar{y}_22 - bar{y}_12 + bar{y}_21 : 1.42
    
```

## 11.5 Modellevaluation

Die inferenzstatistische Evaluation von Haupteffekt- und Interaktionsparametern in ZVA-Modellen in allgemeiner Form kann aus mindestens drei Blickwinkeln betrachtet werden. Analog zur EVA können auch für die ZVA Quadratsummenzerlegungen entwickelt werden und zur Definition entsprechender F-Statistiken eingesetzt werden. Wir folgen hier dem alternativen Ansatz, F-Statistiken für Haupteffekte und Interaktionen vor dem Hintergrund von Modellvergleichen in den entsprechenden ZVA-Modellen zu formulieren. Einen integrativen Blick auf beide Ansätze bietet die Theorie der Allgemeinen Linearen Hypothese, auf deren Diskussion wir hier aber verzichten.

Dementsprechend beschränken wir uns hier auf (1) die Evaluation der Haupteffekte im additiven Modell der  $2 \times 2$  ZVA mit Referenzgruppe und (2) die Evaluation der Interaktion im Modell der  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe. Für beide Evaluationen formulieren wir (1) die zugrundeliegenden Modelle und Teststatistiken, (2) die Testhypothesen und Tests und (3) die Prozeduren zur Testumfangkontrolle und Auswertung des p-Wertes.

### Evaluation von Haupteffekten im additiven Modell der $2 \times 2$ ZVA

**Theorem 11.3** (Teststatistiken für Haupteffekte). *Gegeben sei die Designmatrixform des additiven Modells der  $2 \times 2$  ZVA mit Referenzgruppe*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (11.51)$$

wobei die Spalten von  $X$  bezeichnet seien durch

$$X := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2}) \in \mathbb{R}^{n \times 3}. \quad (11.52)$$

Dann gelten

(A) Eine F-Teststatistik für den Haupteffekt von Faktor A ist die F-Statistik unter der Partitionierung

$$X := (X_{\mu_0} \quad X_{\beta_2} \quad X_{\alpha_2}), \beta := \begin{pmatrix} \mu_0 \\ \beta_2 \\ \alpha_2 \end{pmatrix} \quad \text{und} \quad X_0 := (X_{\mu_0} \quad X_{\beta_2}), \beta_0 := \begin{pmatrix} \mu_0 \\ \beta_2 \end{pmatrix}. \quad (11.53)$$

(B) Eine F-Teststatistik für den Haupteffekt von Faktor B ist die F-Statistik unter der Partitionierung

$$X := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2}), \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \quad \text{und} \quad X_0 := (X_{\mu_0} \quad X_{\alpha_2}), \beta_0 := \begin{pmatrix} \mu_0 \\ \alpha_2 \end{pmatrix}. \quad (11.54)$$

◦

**Definition 11.5** (Testhypothesen und Tests für Haupteffekte). Gegeben sei das Modell der additiven  $2 \times 2$  ZVA mit Referenzgruppe. Die F-Teststatistiken für die Haupteffekte von Faktor A und B seien mit  $F_A$  und  $F_B$  bezeichnet und wie oben definiert. Dann gilt:

(A) Der kritische-Wert-basierte Test

$$\phi_A(y) := 1_{\{F_A \geq k\}} \text{ mit Nullhypothese } H_0^A : \alpha_2 = 0 \quad (11.55)$$

definiert den *F-Test des Haupteffekts von Faktor A*.

(B) Der kritische-Wert-basierte Test

$$\phi_B(y) := 1_{\{F_B \geq k\}} \text{ mit Nullhypothese } H_0^B : \beta_2 = 0 \quad (11.56)$$

definiert den *F-Test des Haupteffekts von Faktor B*.

•

**Theorem 11.4** (Testumfangkontrolle und p-Werte für Haupteffekte). *Mit obigen Definitionen und der KVF  $\varphi(\cdot; n_1, n_2)$  der  $f$ -Verteilung gelten:*

(A)  $\phi_A$  ist ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$ , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0}^A := \varphi^{-1}(1 - \alpha_0; 1, n - 3). \tag{11.57}$$

Der zu einem beobachteten Wert  $f_A$  von  $F_A$  assoziierte  $p$ -Wert ist gegeben durch

$$p\text{-Wert} := 1 - \varphi(f_A; 1, n - 3). \tag{11.58}$$

(B)  $\phi_B$  ist ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$ , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0}^B := \varphi^{-1}(1 - \alpha_0; 1, n - 3). \tag{11.59}$$

Der zu einem beobachteten Wert  $f_B$  von  $F_B$  assoziierte  $p$ -Wert ist gegeben durch

$$p\text{-Wert} := 1 - \varphi(f_B; 1, n - 3). \tag{11.60}$$

◦

### Anwendungsbeispiel

Die Anwendung des oben formulierten Evaluationsverfahrens für die Haupteffekte im Modell der additiven  $2 \times 2$  ZVA mit Referenzgruppe anhand des Beispieldatensatzes demonstriert folgender **R**-Code.

```
D = read.csv("data/11_Zweifaktorielle_Varianzanalyse.csv") # Datensatz
A1B1 = D$dBDI[D$Setting == "F2F" & D$Variant == "MND" ] # face-to-face, mindfulness
A1B2 = D$dBDI[D$Setting == "F2F" & D$Variant == "EXC" ] # face-to-face, exercise
A2B1 = D$dBDI[D$Setting == "ONL" & D$Variant == "MND" ] # online, mindfulness
A2B2 = D$dBDI[D$Setting == "ONL" & D$Variant == "EXC" ] # online, exercise
I = 2 # Anzahl Level Faktor Setting
J = 2 # Anzahl Level Faktor Variante
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
X = kronecker(D,C) # Kroneckerprodukt-Designmatrix
XH = list(X[,c(1,3,2)], X) # Modellvarianten
alpha_0 = 0.05 # Signifikanzlevel
Eff = rep(NaN,2) # F-Teststatistik Arrayinitialisierung
k_alpha_0 = rep(NaN,2) # Kritischer Wert Arrayinitialisierung
phi = rep(NaN,2) # Testwert Arrayinitialisierung
p_vals = rep(NaN,2) # p-Wert Arrayinitialisierung
for(i in 1:2){ # Iteration über Modellvarianten
  X = XH[[i]] # Designmatrix vollständiges Modell
  X_0 = X[,-3] # Designmatrix reduziertes Modell
  p = ncol(X) # Anzahl Parameter vollständiges Modell
  p_0 = ncol(X_0) # Anzahl Parameter reduziertes Modell
  p_1 = p - p_0 # Anzahl zusätzlicher Parameter im vollst. Modell
  beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y # Betaparameterschätzer reduziertes Modell
  beta_hat = solve(t(X)%*%X)%*%t(X)%*%y # Betaparameterschätzer vollständiges Modell
  eps_hat_0 = y - X_0%*%beta_hat_0 # Residuenvektor reduziertes Modell
  eps_hat = y - X%*%beta_hat # Residuenvektor vollständiges Modell
  eh0_eh0 = t(eps_hat_0)%*%eps_hat_0 # RQS reduziertes Modell
  eh_eh = t(eps_hat)%*%eps_hat # RQS vollständiges Modell
  sigsq_hat = eh_eh/(n-p) # Varianzparameterschätzer vollst. Modell
  Eff[i] = ((eh0_eh0-eh_eh)/p_1)/sigsq_hat # F-Statistik
  k_alpha_0[i] = qf(1-alpha_0, p_1, n-p) # kritischer Wert
  if(Eff[i] >= k_alpha_0[i]){ phi[i] = 1 } # H_0 ablehnen
  else { phi[i] = 0 } # H_0 nicht ablehnen
  p_vals[i] = 1 - pf(Eff[i], p_1,n-p) # p-Wert
}
```

```

      f      k phi p.Wert
Setting 0.172 4.057 0 0.68
Variant 51.356 4.057 1 0.00
    
```

Für den Haupteffekt des Faktors *Setting* ergibt sich dabei ein Wert der  $F$ -Statistik von  $F_A = 0.17$  bei einem kritischen Wert von  $k_{0.05} = 4.06$ . Die Nullhypothese eines wahren, aber unbekanntem Effektparameterwertes von Null für den Faktor *Setting* würde im Lichte der in Abbildung 11.4 visualisierten Daten also nicht verworfen werden.

Für den Haupteffekt des Faktors *Variant* dagegen ergibt ein Wert der  $F$ -Statistik von  $F_B = 51.36$  bei dem gleichen kritischen Wert von  $k_{0.05} = 4.06$ . Die Nullhypothese eines wahren, aber unbekanntem Effektparameterwertes von Null für den Faktor *Variant* würde im Lichte der in Abbildung 11.4 visualisierten Daten mit einem Signifikanzlevel von  $\alpha_0 := 0.05$  also verworfen werden.

Folgender **R**-Code demonstriert die gleiche Analyse mithilfe der **R**-Funktion `aov()`.

```

D      = read.csv("_data/11_Zweifaktorielle_Varianzanalyse.csv")      # Datensatz
res.aov = aov(dBDI ~ Setting + Variant, data = D)                  # Modellformulierung und Modellschätzung
summary(res.aov)                                                  # Modellevaluation

      Df Sum Sq Mean Sq F value Pr(>F)
Setting  1  1.02   1.02  0.172   0.68
Variant  1 305.02  305.02 51.356 5.78e-09 ***
Residuals 45 267.27   5.94
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

Folgender **R**-Code schließlich demonstriert die gleiche Analyse mithilfe der **R**-Funktionen `lm()` und `anova()`.

```

D      = read.csv("_data/11_Zweifaktorielle_Varianzanalyse.csv")      # Datensatz
glm    = lm(dBDI ~ Setting + Variant, data = D)                  # Modellformulierung und Modellschätzung
anova(glm)                                                       # Modellevaluation

Analysis of Variance Table

Response: dBDI
      Df Sum Sq Mean Sq F value Pr(>F)
Setting  1  1.021   1.021  0.1719  0.6804
Variant  1 305.021  305.021 51.3559 5.779e-09 ***
Residuals 45 267.271   5.939
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

### Evaluation der Interaktion im Modell der $2 \times 2$ ZVA mit Interaktion

**Theorem 11.5** (Teststatistik für die Interaktion). *Gegeben sei die Designmatrixform des additiven Modells der  $2 \times 2$  ZVA mit Interaktion Referenzgruppe*

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n), \tag{11.61}$$

wobei die Spalten von  $X$  bezeichnet seien durch

$$X := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2} \quad X_{\gamma_{22}}) \in \mathbb{R}^{n \times 4}. \tag{11.62}$$

Dann gilt: Eine F-Teststatistik für die Interaktion von Faktor A und B ist die F-Statistik unter der Partitionierung

$$X := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2} \quad X_{\gamma_{22}}), \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix}, p := 4 \quad \text{und} \tag{11.63}$$

$$X_0 := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2}), \beta_0 := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix}, p_0 := 3.$$

◦

**Definition 11.6** (Testhypothese und Test für die Interaktion). 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

$$\phi_{A \times B}(y) := 1_{\{F_{A \times B} \geq k\}} \text{ mit Nullhypothese } H_0^{A \times B} : \gamma_{22} = 0 \tag{11.64}$$

den F-Test der Interaktion von Faktor A und Faktor B.

•

**Theorem 11.6** (Testumfangkontrolle und p-Wert für die Interaktion). Mit der obigen Definition und der KVF  $\varphi(\cdot; n_1, n_2)$  der f-Verteilung gilt, dass  $\phi_{A \times B}$  ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$  ist, wenn der kritische Wert definiert ist durch

$$k_{\alpha_0}^{A \times B} := \varphi^{-1}(1 - \alpha_0; 1, n - 4). \tag{11.65}$$

Der zu einem beobachteten Wert  $f_{A \times B}$  von  $F_{A \times B}$  assoziierte p-Wert ist gegeben durch

$$p\text{-Wert} := 1 - \varphi(f_{A \times B}; 1, n - 4). \tag{11.66}$$

◦

### Anwendungsbeispiel

Die Anwendung des oben formulierten Evaluationsverfahrens für die Interaktion der Faktoren im Modell der  $2 \times 2$  ZVA mit Interaktion und Referenzgruppe anhand des Beispieldatensatzes demonstriert folgender R-Code.

```
D = read.csv("_data/11_Zweifaktorielle_Varianzanalyse.csv") # Datensatz
A1B1 = D$dBDI[D$Setting == "F2F" & D$Variant == "MND" ] # face-to-face, mindfulness
A1B2 = D$dBDI[D$Setting == "F2F" & D$Variant == "EXC" ] # face-to-face, exercise
A2B1 = D$dBDI[D$Setting == "ONL" & D$Variant == "MND" ] # online, mindfulness
A2B2 = D$dBDI[D$Setting == "ONL" & D$Variant == "EXC" ] # online, exercise
I = 2 # Anzahl Level Faktor Setting
J = 2 # Anzahl Level Faktor Variante
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), nrow=I*J, byrow=TRUE)# prototypische Designmatrix
C = matrix(rep(1,n_ij), nrow = n_ij) # prototypischer Zellenvektor
X = kronecker(D,C) # Kroneckerprodukt-Designmatrix
XH = list(X[,c(1,3,2)], X) # Modellvarianten
alpha_0 = 0.05 # Signifikanzlevel
X = XH[[i]] # Designmatrix vollständiges Modell
X_0 = X[, -4] # Designmatrix reduziertes Modell
p = ncol(X) # Anzahl Parameter vollständiges Modell
```

```

p_0      = ncol(X_0)                    # Anzahl Parameter reduziertes Modell
p_1      = p - p_0                    # Anzahl zusätzlicher Parameter im vollst. Modell
beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y # Betaparameterschätzer reduziertes Modell
beta_hat  = solve(t(X) %*%X )%*%t(X) %*%y    # Betaparameterschätzer vollständiges Modell
eps_hat_0 = y-X_0%*%beta_hat_0          # Residuenvektor reduziertes Modell
eps_hat   = y - X%*%beta_hat           # Residuenvektor vollständiges Modell
eh0_eh0   = t(eps_hat_0) %*% eps_hat_0    # RQS reduziertes Modell
eh_eh     = t(eps_hat) %*% eps_hat       # RQS vollständiges Modell
sigsqr_hat = eh_eh/(n-p)                # Varianzparameterschätzer vollst. Modell
f         = ((eh0_eh0-eh_eh)/p_1)/sigsqr_hat # F-Statistik
k_alpha_0 = qf(1-alpha_0, p_1, n-p)      # kritischer Wert
if(f >= k_alpha_0){phi = 1} else {phi = 0} # Test
p_val     = 1 - pf(f, p_1,n-p)          # p-Wert

```

```

                f      k phi p.Wert
Setting x Variant 1.014 4.062  0  0.319

```

Es ergibt sich dabei ein Wert der  $F$ -Statistik von  $F_{A \times B} = 1.01$  bei einem kritischen Wert von  $k_{0.05} = 4.06$ . Die Nullhypothese eines wahren, aber unbekanntem Effektparameterwertes von Null für den Interaktionseffekt würde im Lichte der in Abbildung 11.4 visualisierten Daten also nicht verworfen werden.

Folgender **R**-Code demonstriert die gleiche Analyse mithilfe der **R**-Funktion `aov()`.

```

D      = read.csv("_data/11_Zweifaktorielle_Varianzanalyse.csv") # Datensatz
res.aov = aov(dBDI ~ Setting + Variant + Setting:Variant, data = D) # Modellformulierung und Modellschätzung
summary(res.aov) # Modellevaluation

```

```

                Df Sum Sq Mean Sq F value Pr(>F)
Setting         1  1.02   1.02  0.172  0.680
Variant         1 305.02 305.02 51.372 6.5e-09 ***
Setting:Variant 1   6.02   6.02  1.014  0.319
Residuals      44 261.25   5.94
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Folgender **R**-Code schließlich demonstriert die gleiche Analyse mithilfe der **R**-Funktionen `lm` und `anova()`.

```

D      = read.csv("_data/11_Zweifaktorielle_Varianzanalyse.csv") # Datensatz
glm    = lm(dBDI ~ Setting + Variant + Setting:Variant, data = D) # Modellformulierung und Modellschätzung
anova(glm) # Modellevaluation

```

Analysis of Variance Table

```

Response: dBDI
                Df Sum Sq Mean Sq F value Pr(>F)
Setting         1  1.021   1.021  0.1719  0.6804
Variant         1 305.021 305.021 51.3719 6.502e-09 ***
Setting:Variant 1   6.021   6.021  1.0140  0.3194
Residuals      44 261.250   5.938
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## 11.6 Literaturhinweise

Die Popularität varianzanalytischer Verfahren wird im Allgemeinen auf Fisher (1925a) und Fisher (1935) zurückgeführt. Everitt und Howell (2005) und Stigler (1986) geben einen kurzen bzw. einen ausführlichen historischen Überblick.

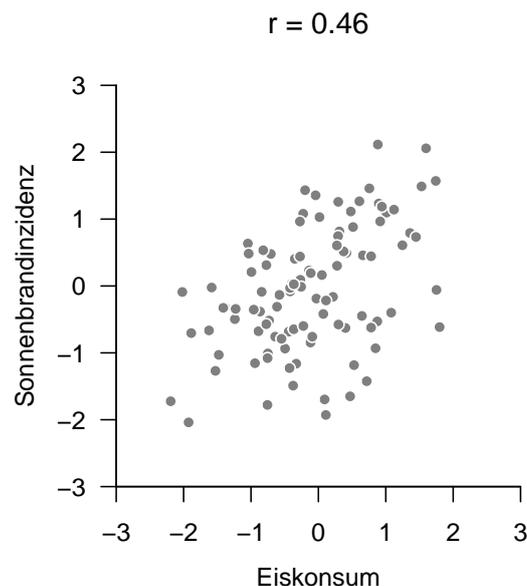
## 11.7 Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario der zweifaktoriellen Varianzanalyse (ZVA).
2. Aus wie vielen Datenpunkten besteht ein Datensatz eines  $3 \times 4$  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  $\mu_0$ ,  $\alpha_2$  und  $\beta_2$  im additiven Modell der ZVA mit RG.
7. Bestimmen Sie  $\mu_{ij}$  für  $\mu_0 := 2$ ,  $\alpha_2 := -1$ ,  $\beta_2 := 3$  im additiven Modell der ZVA mit RG.
8. Geben Sie die Designmatrixform des Modells einer additiven  $2 \times 2$  ZVA mit RG für  $n_{ij} := 1$  an.
9. Geben Sie die Designmatrixform des Modells einer additiven  $2 \times 2$  ZVA mit RG für  $n_{ij} := 3$  an.
10. Geben Sie die Definition des Modells der ZVA mit Interaktion und Referenzgruppe wieder.
11. Erläutern Sie die Bedeutung der Parameter  $\mu_0$ ,  $\alpha_2$ ,  $\beta_2$  und  $\gamma_{22}$  im Modell der ZVA mit Interaktion und RG.
12. Bestimmen Sie  $\mu_{ij}$  für  $\mu_0 := 2$ ,  $\alpha_2 := 1$ ,  $\beta_2 := -1$ ,  $\gamma_{22} := 3$  im Modell der ZVA mit Interaktion und RG.
13. Geben Sie die Designmatrixform des Modells einer  $2 \times 2$  ZVA mit Interaktion und RG für  $n_{ij} := 1$  an.
14. Geben Sie die Designmatrixform des Modells einer  $2 \times 2$  ZVA mit Interaktion und RG für  $n_{ij} := 3$  an.
15. Geben Sie das Theorem zur Betaparameterschätzung im additiven  $2 \times 2$  ZVA-Modell mit RG wieder.
16. Geben Sie das Theorem zur Betaparameterschätzung im  $2 \times 2$  ZVA-Modell mit Interaktion und RG wieder.

# 12 Partielle Korrelation

## 12.1 Motivation

Zur Motivation des Begriffs der partiellen Korrelation betrachten wir zunächst den in Abbildung 12.1 visualisierten Beispieldatensatz zum Zusammenhang von Eiskonsum und Sonnenbrandinzidenz. Wir stellen uns vor, dass jeder der in Abbildung 12.1 abgebildeten Datenpunkte ein Wertepaar aus einem durchschnittlichen und normalisierten Eiskonsum und einer durchschnittlichen und normalisierten Sonnenbrandinzidenz eines Landes über einen gewissen Erhebungszeitraum ist. Visuell betrachtet lässt sich eine Tendenz erkennen, dass hohe Werte des Eiskonsums mit eher hohen Werten der Sonnenbrandinzidenz auftreten, während niedrige Werte des Eiskonsums mit eher niedrigen Werten der Sonnenbrandinzidenz zusammen auftreten. Die Bestimmung des Stichprobenkorrelationskoeffizienten zu diesem Datensatz ergibt mit  $r = 0.46$  eine mittelstarke positive Korrelation.



**Abbildung 12.1** Beispielszenario zur Evaluation bedingter und partieller Korrelationen.

Intuitiv ist es jedoch eher unplausibel, dass Eiskonsum ursächlich Sonnenbrand hervorruft bzw. dass Sonnenbrand den Eiskonsum erhöht (allerdings sind diese Szenarien auch nicht gänzlich auszuschließen: ein bestimmter Eiskonsum könnte eine allergische Reaktion hervorrufen mit Symptomen, die dem Sonnenbrand sehr ähnlich sind; andersherum wäre es denkbar, dass bei Sonnenbrand zur Abkühlung gerne Eis konsumiert wird; wir wollen diese eher unplausiblen Erklärungsansätze hier jedoch nicht weiter verfolgen). Der in Abbildung 12.1 dargestellte Datensatz ist also ein Beispiel dafür, dass Korrelation als Maß für den linearen Zusammenhang zweier Zufallsvariablen lediglich ein Maß für die Koinzidenz

bestimmter Datenwerte ist, jedoch keine Kausalerklärung der Werte einer abhängigen Variable aus den Werten einer unabhängigen Variable impliziert. In Kurzform hat sich zur Beschreibung dieser Tatsache seit Beginn der modernen Inferenzstatistik am Anfang des 20. Jahrhunderts der Leitsatz “*correlation is not causation*” eingebürgert.

Basierend auf dem negativen Ergebnis, dass eine mittelstarke Korrelation wie im Beispiel von Eiskonsum und Sonnenbrandinzidenz nur sehr unplausibel durch eine direkte kausale Beziehung der beiden Variablen zu erklären ist, stellt sich die Frage, inwieweit andere datenanalytische Verfahren hier Abhilfe schaffen können. Dabei stellt sich natürlich zunächst das philosophische Problem, was Kausalität eigentlich bedeuten soll, und als nächstes die Frage, wie ein solcher Begriff mit den Mitteln der Wahrscheinlichkeitstheorie und Inferenzstatistik evaluiert werden könnte. Diesen Ansatz verfolgt das Gebiet der *Kausalen Inferenz*, repräsentiert z.B. durch die Arbeiten von Pearl (2000) und Imbens und Rubin (2015) repräsentiert. Wir wollen an dieser Stelle diesen Ansatz nicht vertiefen, sondern stattdessen fragen, wie im obigen Beispiel anhand weiterer Daten die beobachtete Korrelation von Eiskonsum und Sonnenbrandinzidenz so aufgeklärt werden kann, dass die statistische Beschreibung des in Abbildung 12.1 dargestellten Datensatzes plausibler erscheint. Dies ist das zentrale Thema der partiellen Korrelation.

Dazu nehmen wir an, dass der Zusammenhang von Eiskonsum und Sonnenbrandinzidenz (Abbildung 12.2 A) plausibel durch die Kovariation beider Variablen mit einer dritten Variable, nämlich der Anzahl der im Erhebungszeitraum und Land auftretenden Anzahl an Sommertagen, d.h. Tagen mit einer maximalen Temperatur von über 25° Celsius, erklärt werden kann (Abbildung 12.2 B).

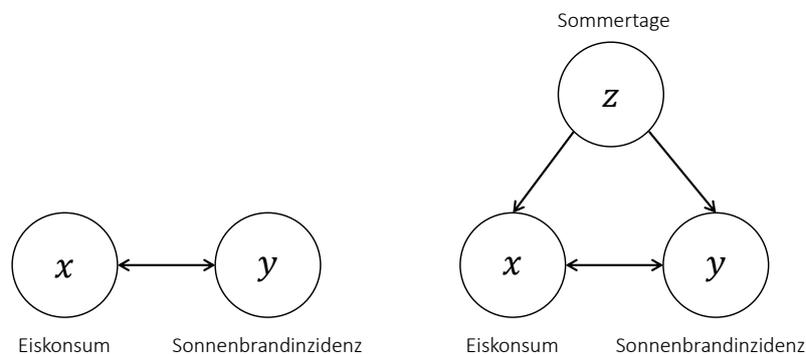


Abbildung 12.2 Zufallsvariablen im Beispielszenario.

Intuitiv erklärt sich die positive Korrelation von Eiskonsum und Sonnenbrandinzidenz dann wie folgt: Treten im Erhebungszeitraum in einem Land mehr Sommertage auf, so steigt in diesem Land sowohl der Eiskonsum als auch die Sonnenbrandinzidenz, treten dagegen weniger Sommertage auf, so fallen in diesem Land sowohl der Eiskonsum als auch die Sonnenbrandinzidenz. Lässt man die Anzahl der Sommertage außer Acht, so treten also hohe Werte von Eiskonsum und Sonnenbrandinzidenz als auch niedrige Werte von Eiskonsum und Sonnenbrandinzidenz häufig zusammen auf und es ergibt sich die in Abbildung 12.1 implizierte positive Korrelation.

Die entscheidende Frage in diesem Kontext ist also, ob bei gleicher Anzahl von Sommertagen Evidenz für eine Korrelation von Eiskonsum und Sonnenbrandinzidenz besteht oder nicht. In diesem Fall würde die Kovarianz von Eiskonsum und Sonnenbrandinzidenz also

bedingt auf einer konstanten Anzahl von Sommertagen betrachtet werden. Die datenanalytischen Werkzeuge, bei Vorliegen von Realisationen von drei Zufallsvariablen eben diese Form eines bedingten Zusammenhangs zu evaluieren, stellen der Begriff der *bedingten Korrelation* und der eng verwandte Begriff der *partiellen Korrelation* bereit. Intuitiv handelt es sich dabei um die Korrelation zweier Zufallsvariablen (z.B. Eiskonsum und Sonnenbrandinzidenz), nachdem aus beiden Zufallsvariablen der Einfluß einer dritten Zufallsvariable (z.B. Anzahl an Sommertagen) “herausgerechnet” wurde. Die Begriffe der bedingten und partiellen Korrelation sind dabei nicht auf das Szenario von drei Zufallsvariablen beschränkt, sondern können für beliebig viele Zufallsvariablen generalisiert werden. Wir beschränken uns in diesem Abschnitt allerdings auf das Szenario dreier Zufallsvariablen, um die Grundlagen der Theorie zu verdeutlichen.

Wir gehen dabei wie folgt vor: In Abschnitt 12.2 führen wir mit der bedingten Kovarianz und der bedingten Korrelation zunächst allgemeine Maße für den auf den Werten einer dritten Zufallsvariable bedingten linear-affinen Zusammenhang zweier Zufallsvariablen ein, verdeutlichen dann die Begriffe anhand des Szenarios dreier gemeinsam multivariat normalverteilter Zufallsvariablen und diskutieren schließlich den Zusammenhang zwischen bedingter Korrelation und paarweisen (unbedingten) Korrelation. In Abschnitt 12.3 führen wir mit der partiellen Korrelation dann ein regressionsbasiertes Maß für den bedingten Zusammenhang zweier Zufallsvariablen ein. Dabei ergibt sich insbesondere, dass im Falle von gemeinsam multivariat normalverteilten Zufallsvariablen bedingte und partielle Korrelation identisch sind. Wir schließen diesen Abschnitt mit der Evaluation der partiellen Korrelation von Eiskonsum und Sonnenbrandinzidenz im Lichte des Wissens um die Anzahl an Sommertagen für den in Abbildung 12.1 visualisierten Beispieldatensatz.

## 12.2 Bedingte Korrelation

Wir definieren zunächst die bedingte Kovarianz und die bedingte Korrelation zweier Zufallsvariablen, gegeben eine dritte Zufallsvariable.

**Definition 12.1** (Bedingte Kovarianz und bedingte Korrelation). Gegeben seien drei Zufallsvariablen  $x, y, z$  einer gemeinsamen Verteilung  $\mathbb{P}(x, y, z)$ .

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

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

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

$$\rho(x, y|z) := \frac{\mathbb{C}(x, y|z)}{\mathbb{S}(x|z)\mathbb{S}(y|z)} \quad (12.1)$$

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

•

Die bedingte Kovarianz zweier Zufallsvariablen ist also definiert als die Kovarianz zweier Zufallsvariablen in einer auf einer dritten Zufallsvariable bedingten Verteilung. Gleiches gilt für die bedingte Korrelation zweier Zufallsvariablen. Durch Vertauschen in obiger Definition kann man analog  $\rho(y, z|x)$  und  $\rho(x, z|y)$  definieren. Wir verdeutlichen Definition 12.1 als nächstes an einem Beispiel.

### Beispiel

Die Zufallsvariablen  $x, y, z$  seien multivariat normalverteilt, d.h.

$$v \sim N(\mu, \Sigma) \quad (12.2)$$

mit

$$\mu := \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 & \sigma_{x,z}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 & \sigma_{y,z}^2 \\ \sigma_{z,x}^2 & \sigma_{z,y}^2 & \sigma_z^2 \end{pmatrix}. \quad (12.3)$$

Wir nehmen an, dass wir die bedingte Korrelation von  $x$  und  $y$  gegeben  $z$  bestimmen wollen und wenden uns entsprechend der bedingten Verteilung von  $x$  und  $y$  gegeben  $z$  zu. Nach Theorem 4.7 wissen wir, dass diese bedingte Verteilung ebenfalls eine Normalverteilung ist, deren Kovarianzmatrixparameter wir aus dem Kovarianzmatrixparameter der gemeinsamen Verteilung von  $x, y, z$  bestimmen können. Wir definieren zu diesem Zweck zunächst

$$\Sigma_{x,y} := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 \end{pmatrix}, \quad \Sigma_z := (\sigma_z^2) \quad \text{und} \quad \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^T := \begin{pmatrix} \sigma_{x,z}^2 \\ \sigma_{y,z}^2 \end{pmatrix}, \quad (12.4)$$

so dass für den Kovarianzmatrixparameter der gemeinsamen Verteilung von  $x, y, z$  gilt, dass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix}. \quad (12.5)$$

Mit Theorem 4.7 ist dann die Kovarianzmatrix des Zufallsvektors  $(x, y)^T$  gegeben durch

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z} \Sigma_z^{-1} \Sigma_{z,(x,y)}. \quad (12.6)$$

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

$$\Sigma_{x,y|z} = \begin{pmatrix} \mathbb{C}(x, x|z) & \mathbb{C}(x, y|z) \\ \mathbb{C}(y, x|z) & \mathbb{C}(y, y|z) \end{pmatrix}. \quad (12.7)$$

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

$$\rho(x, y|z) = \frac{\mathbb{C}(x, y|z)}{\sqrt{\mathbb{C}(x, x|z)} \sqrt{\mathbb{C}(y, y|z)}}. \quad (12.8)$$

Sei konkret etwa der Kovarianzmatrixparameter von  $v = (x, y, z)^T$  gegeben als

$$\Sigma := \begin{pmatrix} 1.0 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.5 \\ 0.9 & 0.5 & 1.0 \end{pmatrix}. \quad (12.9)$$

Dann ergibt sich

$$\rho(x, y) = 0.50 \text{ und } \rho(x, y|z) \approx 0.13. \quad (12.10)$$

Folgender **R**-Code demonstriert die Auswertung dieser bedingten Korrelation.

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

```
rho(x,y) : 0.5
rho(x,y|z) : 0.13
```

## Bedingte Korrelation bei Normalverteilung

Für den Fall dreier gemeinsam normalverteilter Zufallsvariablen eröffnet folgendes Theorem eine Möglichkeit, die bedingte Korrelation zweier dieser Zufallsvariablen, gegeben die dritte auf Grundlage der (unbedingten) paarweisen Korrelationen der Zufallsvariablen zu bestimmen. So kann bei gemeinsamer Normalverteilung von  $x, y, z$  zum Beispiel  $\rho(x, y|z)$  aus den Korrelationen  $\rho(x, y)$ ,  $\rho(x, z)$ , und  $\rho(y, z)$  bestimmt werden. Speziell gilt folgendes Theorem.

**Theorem 12.1** (Bedingte Korrelation und Korrelationen bei Normalverteilung).  *$x, y, z$  seien drei gemeinsam multivariat normalverteilte Zufallsvariablen. Dann gilt:*

$$\rho(x, y|z) = \frac{\rho(x, y) - \rho(x, z)\rho(y, z)}{\sqrt{1 - \rho(x, z)^2}\sqrt{1 - \rho(y, z)^2}}. \quad (12.11)$$

◦

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

$$\Sigma := \begin{pmatrix} 1 & \rho(x, y) & \rho(x, z) \\ \rho(y, x) & 1 & \rho(y, z) \\ \rho(z, x) & \rho(z, y) & 1 \end{pmatrix}. \quad (12.12)$$

Wir definieren nun zunächst

$$\Sigma_{x,y} := \begin{pmatrix} 1 & \rho(x, y) \\ \rho(y, x) & 1 \end{pmatrix}, \Sigma_z := (1) \text{ und } \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^T := \begin{pmatrix} \rho(x, z) \\ \rho(y, z) \end{pmatrix}, \quad (12.13)$$

sodass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix}. \quad (12.14)$$

Mit dem Theorem zu bedingten Normalverteilungen (siehe Theorem 4.7) ist dann die Kovarianzmatrix des Zufallsvektors  $(x, y)^T$  gegeben durch

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z} \Sigma_z^{-1} \Sigma_{z,(x,y)}. \quad (12.15)$$

Es ergibt sich also

$$\begin{aligned} \begin{pmatrix} \sigma_{x,x|z}^2 & \sigma_{x,y|z}^2 \\ \sigma_{y,x|z}^2 & \sigma_{y,y|z}^2 \end{pmatrix} &= \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x,z) \\ \rho(y,z) \end{pmatrix} (1)^{-1} \begin{pmatrix} \rho(x,z) & \rho(y,z) \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x,z)\rho(x,z) & \rho(x,z)\rho(y,z) \\ \rho(y,z)\rho(x,z) & \rho(y,z)\rho(y,z) \end{pmatrix} \\ &= \begin{pmatrix} 1 - \rho(x,z)^2 & \rho(x,y) - \rho(x,z)\rho(y,z) \\ \rho(y,x) - \rho(y,z)\rho(x,z) & 1 - \rho(y,z)^2 \end{pmatrix}. \end{aligned} \quad (12.16)$$

Damit folgt dann direkt

$$\rho(x,y|z) = \frac{\sigma_{x,y|z}^2}{\sqrt{\sigma_{x,x|z}^2} \sqrt{\sigma_{y,y|z}^2}} = \frac{\rho(x,y) - \rho(x,z)\rho(y,z)}{\sqrt{1 - \rho(x,z)^2} \sqrt{1 - \rho(y,z)^2}}. \quad (12.17)$$

□

## 12.3 Partielle Korrelation

Wir definieren als nächstes die partielle Korrelation zweier Zufallsvariablen gegeben eine dritte Zufallsvariable.

**Definition 12.2** (Partielle Korrelation).  $x, y, z$  seien Zufallsvariablen mit linear-affinen Abhängigkeiten zwischen  $x$  und  $z$  sowie zwischen  $y$  und  $z$

$$\begin{aligned} x &:= \beta_0^{(x)} + \beta_1^{(x)} z + e^{(x)} \\ y &:= \beta_0^{(y)} + \beta_1^{(y)} z + e^{(y)}, \end{aligned} \quad (12.18)$$

sodass die Residualvariablen gegeben sind als

$$\begin{aligned} e^{(x)} &= x - \beta_0^{(x)} - \beta_1^{(x)} z \\ e^{(y)} &= y - \beta_0^{(y)} - \beta_1^{(y)} z. \end{aligned} \quad (12.19)$$

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

$$\rho(x, y \setminus z) := \rho(e^{(x)}, e^{(y)}). \quad (12.20)$$

•

Intuitiv entsprechen in obiger Definition die Zufallsvariable  $e^{(x)}$  der Zufallsvariable  $x$ , aus der der Einfluss von  $z$  “herausgerechnet” wurde, und die Zufallsvariable  $e^{(y)}$  der Zufallsvariable  $y$ , aus der der Einfluss von  $z$  “herausgerechnet” wurde. Damit entspricht  $\rho(x, y \setminus z)$  dann intuitiv der Korrelation von  $x$  und  $y$ , aus denen jeweils der Einfluss von  $z$  “herausgerechnet” wurde. Wir geben als nächstes einen Schätzer für die partielle Korrelation zweier Zufallsvariablen, gegeben eine dritte Zufallsvariable an.

**Definition 12.3** (Partielle Stichprobenkorrelation).

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

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

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

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

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

•

Für den Fall, dass  $x, y, z$  multivariat normalverteilt sind, gibt folgendes Theorem, auf dessen Beweis wir hier verzichten wollen, den Zusammenhang zwischen bedingter und partieller Korrelation an.

**Theorem 12.2** (Bedingte und Partielle Korrelation bei Normalverteilung).  $x, y, z$  seien drei gemeinsam multivariat normalverteilte Zufallsvariablen. Dann gilt:

$$\rho(x, y|z) = \rho(x, y \setminus z). \quad (12.21)$$

◦

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

$$\Sigma := \begin{pmatrix} 1 & \rho(x, y) & \rho(x, z) \\ \rho(y, x) & 1 & \rho(y, z) \\ \rho(z, x) & \rho(z, y) & 1 \end{pmatrix}. \quad (12.22)$$

Dann definieren wir

$$A := \begin{pmatrix} 1 & 0 & -\beta_1^{(x)} \\ 0 & 1 & -\beta_1^{(y)} \end{pmatrix} \quad \text{und} \quad b := \begin{pmatrix} -\beta_0^{(x)} \\ -\beta_0^{(y)} \end{pmatrix}, \quad (12.23)$$

sodass gilt

$$\begin{pmatrix} e^{(x)} \\ e^{(y)} \end{pmatrix} = Av + b = \begin{pmatrix} 1 & 0 & -\beta_1^{(x)} \\ 0 & 1 & -\beta_1^{(y)} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} -\beta_0^{(x)} \\ -\beta_0^{(y)} \end{pmatrix} = \begin{pmatrix} x - \beta_0^{(x)} - \beta_1^{(x)} z \\ y - \beta_0^{(y)} - \beta_1^{(y)} z \end{pmatrix}. \quad (12.24)$$

Die partielle Korrelation ergibt sich dann als

$$\rho(x, y \setminus z) = \rho(e^{(x)}, e^{(y)}). \quad (12.25)$$

Mit dem Theorem zur linear-affinen Transformation von multivariaten Normalverteilungen (siehe Theorem 4.3) gilt für den Vektor der Residualvariablen

$$v \sim N(\mu, \Sigma) \Rightarrow \begin{pmatrix} e^{(x)} \\ e^{(y)} \end{pmatrix} = Av + b \sim N(A\mu + b, A\Sigma A^T). \quad (12.26)$$

Der Kovarianzmatrixparameter der resultierenden multivariaten Normalverteilung ergibt sich also zu

$$\begin{aligned} A\Sigma A^T &= \begin{pmatrix} 1 & 0 & -\beta_1^{(x)} \\ 0 & 1 & -\beta_1^{(y)} \end{pmatrix} \begin{pmatrix} 1 & \rho(x, y) & \rho(x, z) \\ \rho(y, x) & 1 & \rho(y, z) \\ \rho(z, x) & \rho(z, y) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\beta_1^{(x)} & -\beta_1^{(y)} \end{pmatrix} \\ &= \begin{pmatrix} 1 - \beta_1^{(x)}\rho(z, x) & \rho(x, y) - \beta_1^{(x)}\rho(z, y) & \rho(x, z) - \beta_1^{(x)} \\ \rho(y, x) - \beta_1^{(y)}\rho(z, x) & 1 - \beta_1^{(y)}\rho(z, y) & \rho(y, z) - \beta_1^{(y)} \\ -\beta_1^{(x)} & -\beta_1^{(y)} & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \beta_1^{(x)}\rho(z, x) - \beta_1^{(x)}\rho(x, z) + (\beta_1^{(x)})^2 & \rho(x, y) - \beta_1^{(x)}\rho(z, y) - \beta_1^{(y)}\rho(x, z) + \beta_1^{(x)}\beta_1^{(y)} \\ \rho(y, x) - \beta_1^{(y)}\rho(z, x) - \beta_1^{(x)}\rho(y, z) + \beta_1^{(x)}\beta_1^{(y)} & 1 - \beta_1^{(y)}\rho(z, y) - \beta_1^{(y)}\rho(y, z) + (\beta_1^{(y)})^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 - 2\beta_1^{(x)}\rho(x, z) + (\beta_1^{(x)})^2 & \rho(x, y) - \beta_1^{(x)}\rho(z, y) - \beta_1^{(y)}\rho(x, z) + \beta_1^{(x)}\beta_1^{(y)} \\ \rho(y, x) - \beta_1^{(y)}\rho(z, x) - \beta_1^{(x)}\rho(y, z) + \beta_1^{(x)}\beta_1^{(y)} & 1 - 2\beta_1^{(y)}\rho(y, z) + (\beta_1^{(y)})^2 \end{pmatrix}. \end{aligned} \quad (12.27)$$

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

$$\rho(x, y) = \frac{\mathbb{S}(x)}{\mathbb{S}(y)}\beta_1 \Leftrightarrow \beta_1 = \frac{\mathbb{S}(y)}{\mathbb{S}(x)}\rho(x, y). \quad (12.28)$$

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

$$\hat{\beta}_1 = \frac{c_{xy}}{s_x^2}, \quad r_{xy} = \frac{c_{xy}}{s_x s_y} \Rightarrow \hat{\beta}_1 = \frac{s_y}{s_x} r_{xy}. \quad (12.29)$$

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

$$\beta_1^{(x)} = \frac{1}{1}\rho(x, z) = \rho(x, z) \quad \text{und} \quad \beta_1^{(y)} = \frac{1}{1}\rho(y, z) = \rho(y, z) \quad (12.30)$$

und damit

$$A\Sigma A^T = \begin{pmatrix} 1 - \rho(x, z)^2 & \rho(x, y) - \rho(x, z)\rho(y, z) \\ \rho(x, y) - \rho(x, z)\rho(y, z) & 1 - \rho(y, z)^2 \end{pmatrix} = \begin{pmatrix} \sigma_{x, x \setminus z}^2 & \sigma_{x, y \setminus z}^2 \\ \sigma_{y, x \setminus z}^2 & \sigma_{y, y \setminus z}^2 \end{pmatrix}. \quad (12.31)$$

Schließlich folgt die partielle Korrelation mit

$$\begin{aligned} \rho(x, y \setminus z) &= \rho(e^{(x)}, e^{(y)}) \\ &= \frac{\sigma_{x, y \setminus z}^2}{\sqrt{\sigma_{x, x \setminus z}^2} \sqrt{\sigma_{y, y \setminus z}^2}} \\ &= \frac{\rho(x, y) - \rho(x, z)\rho(y, z)}{\sqrt{1 - \rho(x, z)^2} \sqrt{1 - \rho(y, z)^2}}. \end{aligned} \quad (12.32)$$

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

Man beachte, dass obiges Theorem im Falle dreier multivariat normalverteilter Zufallsvariablen gilt. Im Allgemeinen, also für beliebige Verteilungen der drei Zufallsvariablen gilt die Identität von bedingter und partieller Korrelationen nicht. Weitere Details in diesem Zusammenhang diskutieren zum Beispiel Lawrance (1976) und Baba, Shibata, und Sibuya (2004).

Aus Theorem 12.2 folgt mit Theorem 12.1 dann unmittelbar, dass bei gemeinsamer Normalverteilung von  $x, y, z$  die partielle Korrelation  $\rho(x, y|z)$  ebenso wie die bedingte Korrelation  $\rho(x, y|z)$  basierend auf den (unbedingten) Korrelationen  $\rho(x, y)$ ,  $\rho(x, z)$  und  $\rho(y, z)$  bestimmt werden kann, bzw. im Falle der jeweiligen Stichprobenäquivalente durch diese geschätzt werden kann.

Folgender **R**-Code demonstriert die Auswertung der partiellen Stichprobenkorrelation basierend auf einem simulierten Datensatz dreier multivariat normalverteilter Zufallsvariablen. Dabei bestimmen wir die partielle Korrelation einmal basierend auf den Residualstichprobenkorrelationen wie in Definition 12.3 und einmal basierend auf den paarweisen Stichprobenkorrelationen anhand von Theorem 12.2. Schließlich stellt das **R**-Paket `ppcor` mit `pcor()` eine Funktion zur automatisierten Auswertung partieller Stichprobenkorrelationen bereit, auch deren Anwendung demonstrieren wir untenstehend. Das Resultat ist natürlich in allen drei Fällen identisch.

```
# Modellformulierung und Datenrealisierung
library(MASS) # multivariate Normalverteilung
set.seed(1) # reproduzierbare Daten
S = matrix(c( 1,0.5,0.9, # Kovarianzmatrixparameter \Sigma
            0.5, 1,0.5,
            0.9,0.5, 1), nrow = 3, byrow = TRUE)

n = 1e6 # Anzahl Realisierungen von v := (x,y,z)^T
xyz = mvrnorm(n,rep(0,3),S) # Realisierungen von v := (x,y,z)^T

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

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

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

```
r(x,y) : 0.5
r(x,y/z) aus Residuenkorrelation : 0.13
r(x,y/z) aus Korrelationen : 0.13
r(x,y/z) aus Toolbox : 0.13
```

## Anwendungsbeispiel

Mithilfe oben eingeführten **R**-Codes wenden wir uns nun abschließend dem eingangs diskutierten Beispiel zum Zusammenhang von Eiskonsum und Sonnenbrandinzidenz zu. Wir nehmen an, dass zu jedem Wertepaar von Eiskonsum ( $x_i$ ) und Sonnenbrandinzidenz ( $y_i$ ) der korrespondierende Wert der Anzahl der Sommertage ( $z_i$ ) im betrachteten Erhebungszeitraum und Land verfügbar ist. Dann eröffnet obige Theorie die Möglichkeit, die partielle Korrelation von Eiskonsum und Sonnenbrandinzidenz nach Korrektur für die Anzahl der Sommertage zu bestimmen.

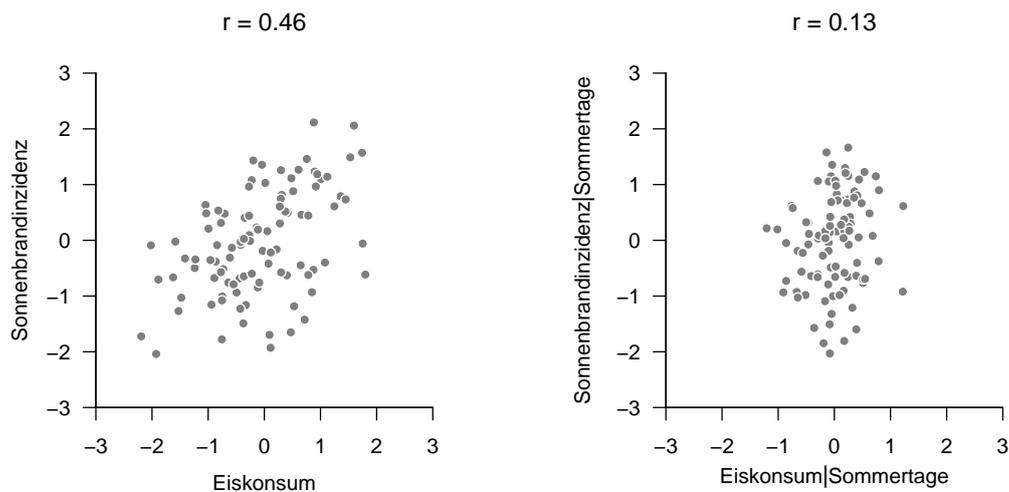


Abbildung 12.3 Evaluation der partiellen Korrelation im Beispielszenario.

Dazu stellt Abbildung 12.3 mit der Achsenbeschriftung **Eiskonsum | Sommertage** die Residualwerte

$$e_i^{(x)} := x_i - \hat{\beta}_0^{(x)} + \hat{\beta}_1^{(x)} z_i \quad (12.33)$$

und mit der Achsenbeschriftung **Sonnenbrandinzidenz | Sommertage** die Residualwerte

$$e_i^{(y)} := y_i - \hat{\beta}_0^{(y)} + \hat{\beta}_1^{(y)} z_i \quad (12.34)$$

dar. Man erkennt, dass kein systematischer Zusammenhang hoher bzw. niedriger Werte von **Eiskonsum | Sommertage** mit hohen bzw. niedrigen Werten von **Sonnenbrandinzidenz | Sommertage** besteht. Die Korrelation dieser Residualwerte beträgt dementsprechend auch nur  $r = 0.17$  und nicht, wie im Falle der nicht für die Kovariation mit der Anzahl der Sommertage korrigierten Werte von Eiskonsum und Sonnenbrandinzidenz,  $r = 0.46$  (vgl. Abbildung 12.1). Der bei der nicht durch die Anzahl der Sommertage informierten Korrelationsanalyse implizierte Zusammenhang von Eiskonsum und Sonnenbrandinzidenz lässt sich also durch die Kovariation beider Variablen mit der Drittvariable Sommertage aufklären bzw. “wegerklären”.

## 12.4 Literaturhinweise

Die Theorie partielle und bedingter Korrelationen findet spätestens seit Beginn der modernen Korrelationsanalyse zu Beginn des 20. Jahrhunderts Beachtung, man vergleiche hierzu zum Beispiel Pearson (1920), Yule (1907) oder Fisher (1924).

## 12.5 Selbstkontrollfragen

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

# 13 Multiple Regression

## 13.1 Anwendungsszenario

Bei der multiplen Regression handelt es sich um die Generalisierung der einfachen linearen Regression auf den Fall von mehr als einer kontinuierlichen unabhängigen Variable. Wie bei der einfachen linearen Regression betrachtet man eine univariate abhängige Variable, deren Werte an randomisierten experimentellen Einheiten bestimmt werden. Die unabhängigen Variablen werden bei der multiplen Regression je nach Kontext *Regressoren*, *Prädiktoren*, *Kovariaten* oder einfach Spalten der Designmatrix genannt. Ähnlich wie bei der einfachen linearen Regression ist das Ziel einer multiplen Regressionsanalyse, das Erklärungspotential für Variation der abhängigen Variablen durch Variationen der unabhängigen Variablen zu quantifizieren. In Erweiterung der einfachen linearen Regression liegt ein Augenmerk dabei insbesondere darauf, den Einfluss einzelner unabhängiger Variablen auf die abhängige Variable im Kontext der Variation anderer unabhängiger Variablen zu ermitteln. Darüber hinaus mag ein Ziel auch die Prädiktion der Werte der abhängigen Variablen für Werte der unabhängigen Variable nach Schätzung der entsprechenden Gewichtungsbetaparameter anhand eines “Trainingsdatensatzes” sein.

### Anwendungsbeispiel

Um einige Grundprinzipien der multiplen Regressionsanalyse zu verdeutlichen, beschränken wir uns in diesem Abschnitt meist auf den Fall zweier kontinuierlicher unabhängiger Variablen. Als Anwendungsbeispiel betrachten wir dazu den Effekt einer Psychotherapie auf die Depressionssymptomatik in Abhängigkeit vom Alter der Patient:innen und der Dauer der Therapie. Tabelle 13.1 zeigt einen Beispieldatensatz, bestehend aus Datenwerten von  $n = 20$  Patient:innen, wobei Alter das Patient:innenalter, Dauer die Therapiedauer und dBDI die BDI-Differenzwerte in der Pre-Post-Interventionsstudie bezeichnen.

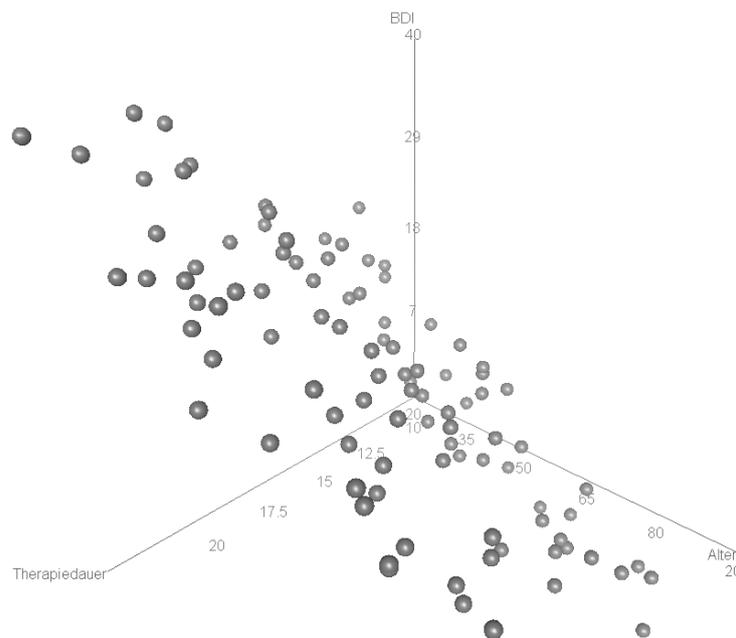
**Tabelle 13.1** BDI-Differenzwerte in Abhängigkeit von Patient:innenalter und Therapiedauer

Alter	Dauer	dBDI
36	23	35.4
42	15	17.5
54	20	17.8
74	14	-5.2
32	15	17.7
74	17	-2.4
77	12	-9.7
60	17	10.2
58	22	19.7
24	16	29.3
32	18	26.3
31	19	26.0
61	18	5.8
43	14	11.0

**Tabelle 13.1** BDI-Differenzwerte in Abhängigkeit von Patient:innenalter und Therapiedauer

Alter	Dauer	dBDI
66	22	15.8
50	20	22.0
63	22	11.2
80	13	-8.8
43	21	28.0
67	17	8.4

Abbildung 13.1 visualisiert den Beispieldatensatz, wobei die dBDI-Werte der Höhe gemäß im dreidimensionalen Raum anhand ihrer Koordinaten über der durch die Variablen Alter und Dauer aufgespannten Ebene abgetragen sind. Die Visualisierung der multiplen Regression, insbesondere für mehr als zwei unabhängige Variablen, stößt sehr schnell an ihre Grenzen.

**Abbildung 13.1** Visualisierung des Beispieldatensatzes als Punktwolke.

## 13.2 Modellformulierung

Das Modell der multiplen Regression ist mit der allgemeinen Form des Allgemeinen Linearen Modells identisch. Der Vollständigkeit halber geben wir folgende Definition.

**Definition 13.1** (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

$$y_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ und } \sigma^2 > 0, \quad (13.1)$$

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* oder *Kovariaten* genannt. Mit

$$x_i := (x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p \text{ und } \beta := (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p \quad (13.2)$$

hat das Modell der multiplen Regression die Datenverteilungsform

$$y_i \sim N(\mu_i, \sigma^2) \text{ u.v. für } i = 1, \dots, n, \text{ wobei } \mu_i := x_i^T \beta. \quad (13.3)$$

In diesem Zusammenhang wird  $x_i \in \mathbb{R}^p$  auch als *iter Merkmalsvektor* bezeichnet. Die Designmatrixform des Modells der multiplen Regression schließlich ist gegeben durch

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (13.4)$$

mit

$$y := (y_1, \dots, y_n)^T, X := (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p} \in \mathbb{R}^{n \times p}, \beta := (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p \text{ und } \sigma^2 > 0. \quad (13.5)$$

Folgender **R**-Code nutzt das Modell der multiplen Regression aus Definition 13.1, um den in Tabelle 13.1 dargestellten Beispieldatensatz zu erzeugen.

```
library(MASS) # multivariate Normalverteilung
set.seed(1) # reproduzierbare Daten
n = 20 # Anzahl Datenpunkte
p = 3 # Anzahl Parameter
x_1 = round(runif(n,20,80)) # Regressorwerte Alter
x_2 = round(runif(n,12,24)) # Regressorwerte Therapiedauer
X = matrix(c(rep(1,n),x_1,x_2), nrow = n) # Designmatrix
I_n = diag(n) # Einheitsmatrix
beta = matrix(c(5,-.5,2), nrow = p) # Betaparametervektor
sigsqr = 10 # Varianzparameter
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y
D = data.frame(Alter = x_1, Dauer = x_2, dBDI = y) # Dataframeformatierung
write.csv(D, "_data/13_Multiple_Regression.csv", row.names = FALSE) # Datenspeicherung
```

### 13.3 Modellschätzung

Das Modell der multiplen Regression erlaubt eine allgemeine Einsicht in die Bedeutung des Betaparameterschätzers

$$\hat{\beta} := (X^T X)^{-1} X^T y \quad (13.6)$$

Insbesondere wird bei der Beschäftigung mit dem Betaparameterschätzer im Kontext der multiplen Regression deutlich, dass  $X^T y$  die Kovariation der Regressoren mit den Daten und  $X^T X$  die Kovariation der Regressoren untereinander quantifiziert, sodass sich für den Betaparameterschätzer eine Intuition als “regressorkovariationsnormalisierte Regressordatenkovariation” ergibt. Wir wollen dies im Folgenden am Beispiel einer multiplen Regression mit einem Interzeptregressor und zwei Regressoren für zwei unabhängige Variablen vertiefen. Dabei zeigen wir einmal die Form des Betaparameterschätzers in Form von Stichprobenkorrelationen (Theorem 13.1) und einmal in Form von partiellen Stichprobenkorrelationen (Theorem 13.2).

**Theorem 13.1** (Betaparameterschätzer und Korrelationen). *Gegeben sei ein multiples Regressionsmodell der Form*

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (13.7)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \\ \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \\ \frac{r_{y,x_2} - r_{y,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_2}} \end{pmatrix}, \quad (13.8)$$

wobei  $\bar{x}$  ein Stichprobenmittel,  $s_x$  eine Stichprobenstandardabweichung und  $r_{x,y}$  eine Stichprobenkorrelation für die entsprechenden  $y_i$ ,  $x_{i1}$  und  $x_{i2}$  mit  $i = 1, \dots, n$  bezeichnet.

◦

*Beweis.* Wir erinnern zunächst daran, dass die Form des Betaparameterschätzers bekanntlich zum System der Normalgleichungen äquivalent ist (vgl. Kapitel 6):

$$\hat{\beta} = (X^T X)^{-1} X^T y \Leftrightarrow X^T X \hat{\beta} = X^T y. \quad (13.9)$$

Ausschreiben des Normalgleichungssystems für den hier betrachteten Spezialfall des ALMs ergibt dann zunächst

$$\begin{aligned} X^T X \hat{\beta} &= X^T y \\ \Leftrightarrow \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \sum_{i=1}^n y_i x_{i2} \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \sum_{i=1}^n y_i x_{i2} \end{pmatrix} \end{aligned}$$

und damit

$$\begin{aligned} X^T X \hat{\beta} &= X^T y \\ \Leftrightarrow \begin{pmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} \\ \hat{\beta}_0 \sum_{i=1}^n x_{i2} + \hat{\beta}_1 \sum_{i=1}^n x_{i1} x_{i2} + \hat{\beta}_2 \sum_{i=1}^n x_{i2}^2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \sum_{i=1}^n y_i x_{i2} \end{pmatrix} \end{aligned}$$

Aus der Gleichung der ersten Vektorkomponenten folgt dann direkt die Form von  $\hat{\beta}_0$  mit

$$\begin{aligned} \sum_{i=1}^n y_i &= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} \\ \Leftrightarrow \frac{1}{n} \sum_{i=1}^n y_i &= \hat{\beta}_0 + \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \frac{1}{n} \sum_{i=1}^n x_{i2} \\ \Leftrightarrow \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \end{aligned} \quad (13.10)$$

Einsetzen dieser Form von  $\hat{\beta}_0$  in die Gleichung der zweiten Vektorkomponenten ergibt dann

$$\begin{aligned} \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} &= \sum_{i=1}^n y_i x_{i1} \\ \Leftrightarrow (\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2) \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} &= \sum_{i=1}^n y_i x_{i1} \\ \Leftrightarrow \bar{y} \sum_{i=1}^n x_{i1} - \hat{\beta}_1 \bar{x}_1 \sum_{i=1}^n x_{i1} - \hat{\beta}_2 \bar{x}_2 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} &= \sum_{i=1}^n y_i x_{i1} \\ \Leftrightarrow \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 - \hat{\beta}_1 \bar{x}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} - \hat{\beta}_2 \bar{x}_2 \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n y_i x_{i1} - \bar{y} \sum_{i=1}^n x_{i1} \\ \Leftrightarrow \hat{\beta}_1 \left( \sum_{i=1}^n x_{i1}^2 - \bar{x}_1 \sum_{i=1}^n x_{i1} \right) + \hat{\beta}_2 \left( \sum_{i=1}^n x_{i1} x_{i2} - \bar{x}_2 \sum_{i=1}^n x_{i1} \right) &= \sum_{i=1}^n y_i x_{i1} - \bar{y} \sum_{i=1}^n x_{i1} \end{aligned}$$

Im Beweis des Theorems zur Ausgleichsgerade (vgl. Kapitel 1) haben wir gesehen, dass

$$\sum_{i=1}^n x_{i1} x_{i1} - \bar{x}_1 \sum_{i=1}^n x_{i1} = \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1), \quad (13.11)$$

$$\sum_{i=1}^n x_{i1} x_{i2} - \bar{x}_2 \sum_{i=1}^n x_{i1} = \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) \text{ und} \quad (13.12)$$

$$\sum_{i=1}^n y_i x_{i1} - \bar{y} \sum_{i=1}^n x_{i1} = \sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1). \quad (13.13)$$

Es ergibt sich also, dass

$$\begin{aligned} \hat{\beta}_1 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1) + \hat{\beta}_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) &= \sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1) \\ \Leftrightarrow \hat{\beta}_1 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1)}{n-1} + \hat{\beta}_2 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{n-1} &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1)}{n-1}. \end{aligned} \quad (13.14)$$

Mit den Definitionen von Stichprobenstandardabweichung und -korrelation folgt dann weiter

$$\begin{aligned} \hat{\beta}_1 s_{x_1} s_{x_1} + \hat{\beta}_2 c_{x_1, x_2} &= c_{y, x_1} \\ \Leftrightarrow \hat{\beta}_1 \frac{s_{x_1} s_{x_1}}{s_y s_{x_1}} + \hat{\beta}_2 \frac{c_{x_1, x_2}}{s_y s_{x_1}} &= \frac{c_{y, x_1}}{s_y s_{x_1}} \\ \Leftrightarrow \hat{\beta}_1 \frac{s_{x_1}}{s_y} + \hat{\beta}_2 \frac{c_{x_1, x_2}}{s_y s_{x_1}} &= r_{y, x_1} \\ \Leftrightarrow \hat{\beta}_1 \frac{s_{x_1}}{s_y} + \hat{\beta}_2 \frac{c_{x_1, x_2} s_{x_2}}{s_y s_{x_1} s_{x_2}} &= r_{y, x_1} \\ \Leftrightarrow \hat{\beta}_1 \frac{s_{x_1}}{s_y} + \hat{\beta}_2 \frac{s_{x_2}}{s_y} r_{x_1, x_2} &= r_{y, x_1} \end{aligned} \quad (13.15)$$

Definition von

$$b_j := \frac{s_{x_j}}{s_y}, \quad j = 1, 2 \quad (13.16)$$

erlaubt dann die Schreibweise

$$b_1 + b_2 r_{x_1, x_2} = r_{y, x_1}. \quad (13.17)$$

Durch Vertauschen der Subskripte folgt analog aus der Gleichung der dritten Vektorkomponenten

$$b_1 r_{x_1, x_2} + b_2 = r_{y, x_2}. \quad (13.18)$$

Insgesamt haben wir also gesehen, dass die Definition des Betaparameterschätzers im vorliegenden ALM-Spezialfall ergibt, dass mit

$$\hat{\beta}_j = b_j \frac{s_y}{s_{x_j}}, \quad j = 1, 2 \quad (13.19)$$

gilt, dass

$$\begin{aligned} r_{y, x_1} &= b_1 + b_2 r_{x_1, x_2} \\ r_{y, x_2} &= b_1 r_{x_1, x_2} + b_2. \end{aligned} \quad (13.20)$$

Damit folgt aus der zweiten Gleichung dann sofort

$$b_2 = r_{y, x_2} - b_1 r_{x_1, x_2}. \quad (13.21)$$

Einsetzen in die erste Gleichung ergibt dann

$$\begin{aligned} b_1 + (r_{y, x_2} - b_1 r_{x_1, x_2}) r_{x_1, x_2} &= r_{y, x_1} \\ \Leftrightarrow b_1 + r_{y, x_2} r_{x_1, x_2} - b_1 r_{x_1, x_2}^2 &= r_{y, x_1} \\ \Leftrightarrow r_{y, x_2} r_{x_1, x_2} + b_1 (1 - r_{x_1, x_2}^2) &= r_{y, x_1} \\ \Leftrightarrow b_1 (1 - r_{x_1, x_2}^2) &= r_{y, x_1} - r_{y, x_2} r_{x_1, x_2} \\ \Leftrightarrow b_1 &= \frac{r_{y, x_1} - r_{y, x_2} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2} \end{aligned} \quad (13.22)$$

Für  $b_2$  ergibt sich damit weiterhin

$$\begin{aligned} b_2 &= r_{y, x_2} - b_1 r_{x_1, x_2} \\ \Leftrightarrow b_2 &= r_{y, x_2} - \left( \frac{r_{y, x_1} - r_{y, x_2} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2} \right) r_{x_1, x_2} \\ \Leftrightarrow b_2 &= \frac{r_{y, x_2} (1 - r_{x_1, x_2}^2)}{1 - r_{x_1, x_2}^2} - \frac{r_{y, x_1} r_{x_1, x_2} - r_{y, x_2} r_{x_1, x_2}^2}{1 - r_{x_1, x_2}^2} \\ \Leftrightarrow b_2 &= \frac{r_{y, x_2} - r_{y, x_2} r_{x_1, x_2}^2 - r_{y, x_1} r_{x_1, x_2} + r_{y, x_2} r_{x_1, x_2}^2}{1 - r_{x_1, x_2}^2} \\ \Leftrightarrow b_2 &= \frac{r_{y, x_2} - r_{y, x_1} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2}. \end{aligned} \quad (13.23)$$

Damit folgen dann aber

$$\begin{aligned} \hat{\beta}_1 &= b_1 \frac{s_y}{s_{x_1}} = \left( \frac{r_{y, x_1} - r_{y, x_2} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2} \right) \frac{s_y}{s_{x_1}} \\ \hat{\beta}_2 &= b_2 \frac{s_y}{s_{x_2}} = \left( \frac{r_{y, x_2} - r_{y, x_1} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2} \right) \frac{s_y}{s_{x_2}} \end{aligned} \quad (13.24)$$

und es ist alles gezeigt.  $\square$

Exemplarisch wollen wir die Stichprobenkorrelationsform des Betaparameterschätzers für  $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{r_{y, x_1} - r_{y, x_2} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2} \frac{s_y}{s_{x_1}}. \quad (13.25)$$

betrachten. Man erkennt unter anderem:

- Nur im Fall  $r_{x_1, x_2} = 0$  und  $s_y = s_{x_1}$  gilt  $\hat{\beta}_1 = r_{y, x_1}$ . Die Betaparameterschätzer der multiplen Regression sind also im Allgemeinen nicht mit den Stichprobenkorrelationen zwischen dem entsprechenden Regressor und den Daten identisch.
- Im Fall  $r_{x_1, x_2} = \pm 1$  ist  $\hat{\beta}_1$  nicht definiert. Vollständig korrelierte Regressoren ergeben also ein nicht schätzbares Modell.
- Je größer  $|r_{x_1, x_2}|$ , desto größer der von  $r_{y, x_1}$  subtrahierte Term  $r_{y, x_2} r_{x_1, x_2}$ . Die paarweise Korrelation von Regressoren untereinander reduziert also den Betaparameterschätzerwert für einen Regressor.
- Je größer  $|r_{y, x_2}|$ , desto größer der von  $r_{y, x_1}$  subtrahierte Term  $r_{y, x_2} r_{x_1, x_2}$ . Neben der paarweisen Korrelation von Regressoren untereinander reduziert also auch eine hohe Stichprobenkorrelation eines anderen Regressors mit den Daten den Wert eines Betaparameterschätzers.
- Bei identischen Korrelationen und gleich bleibender Regressorstandabweichung steigt  $\hat{\beta}_1$  mit  $s_y$ . Intuitiv erhält also ein Regressor einen höheren Betaparameterschätzerwert, wenn er bei gleicher Korrelationsstruktur mehr Datenvariabilität erklärt.

Zusammengefasst verdeutlichen obige Punkte, dass der Wert des Betaparameterschätzers eines Regressors im Modell der multiplen Regression allein im Kontext der Korrelationsstruktur der anderen Regressoren mit sich selbst und den Daten betrachtet werden kann und damit keinen modellunabhängigen Absolutbeitrag eines spezifischen Regressors zur Erklärung der Daten abbildet.

Folgender **R**-Code demonstriert die Äquivalenz der Matrixschreibweise des Betaparameterschätzers und seiner Darstellung durch Stichprobenmittel, Stichprobenkorrelationen und Stichprobenstandardabweichungen.

```
# Daten einlesen und # Modellschätzung
D = read.csv("data/13_Multiple_Regression.csv") # Daten einlesen
y = D$BDI # abhängige Variable
n = length(y) # Anzahl Datenpunkte
X = matrix(c(rep(1,n), D$Alter, D$Dauer), 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 Stichprobenmittel, -standardabweichungen und -korrelationen
y12 = cbind(y, X[, -1]) # Matrix (y, x_1, x_2)
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
```

```
beta_hat ALM-Schätzer : 11.43949 -0.5716188 1.851319
beta_hat Deskriptivstatistiken : 11.43949 -0.5716188 1.851319
```

Folgendes Theorem stellt den Bezug zwischen multipler Regression und partiellen Korrelationen zwischen unabhängigen und abhängigen Variablen für das betrachtete Szenario her.

**Theorem 13.2** (Betaparameterschätzer und partielle Korrelationen). *Gegeben sei ein multiples Regressionsmodell der Form*

$$y = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (13.26)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \\ r_{y,x_1 \setminus x_2} \sqrt{\frac{1-r_{y,x_2}^2}{1-r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}} \\ r_{y,x_2 \setminus x_1} \sqrt{\frac{1-r_{y,x_1}^2}{1-r_{x_2,x_1}^2}} \frac{s_y}{s_{x_2}} \end{pmatrix}, \quad (13.27)$$

wobei für  $1 \leq k, l \leq 2$  und  $i = 1, \dots, n$

- $r_{y,x_k \setminus x_l}$  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.

◦

*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

$$\hat{\beta}_1 = \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}}. \quad (13.28)$$

Weiterhin haben wir bereits gesehen (vgl. Kapitel 12), 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

$$r_{y,x_1 \setminus x_2} = \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \quad (13.29)$$

gegeben ist. Für  $\hat{\beta}_1$  ergibt sich somit

$$\begin{aligned} \hat{\beta}_1 &= \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \\ &\Leftrightarrow (1 - r_{x_1,x_2}^2) \hat{\beta}_1 = (r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}) \frac{s_y}{s_{x_1}} \\ &\Leftrightarrow \frac{1 - r_{x_1,x_2}^2}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \hat{\beta}_1 = \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}} \\ &\Leftrightarrow \frac{1 - r_{x_1,x_2}^2}{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \hat{\beta}_1 = r_{y,x_1 \setminus x_2} \frac{s_y}{s_{x_1}} \end{aligned} \quad (13.30)$$

und damit weiter

$$\begin{aligned} \hat{\beta}_1 &= r_{y,x_1 \setminus x_2} \frac{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}} \\ &\Leftrightarrow \hat{\beta}_1 = r_{y,x_1 \setminus x_2} \frac{\sqrt{1 - r_{y,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}}{\left(\sqrt{1 - r_{x_1,x_2}^2}\right)^2} \frac{s_y}{s_{x_1}} \\ &\Leftrightarrow \hat{\beta}_1 = r_{y,x_1 \setminus x_2} \frac{\sqrt{1 - r_{y,x_2}^2}}{\sqrt{1 - r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}} \\ &\Leftrightarrow \hat{\beta}_1 = r_{y,x_1 \setminus x_2} \sqrt{\frac{1 - r_{y,x_2}^2}{1 - r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}}. \end{aligned} \quad (13.31)$$

□

Wir betrachten exemplarisch

$$\hat{\beta}_1 = r_{y,x_1|x_2} \sqrt{\frac{1 - r_{y,x_2}^2}{1 - r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}}. \quad (13.32)$$

Im Allgemeinen gilt, dass  $\hat{\beta}_1 \neq r_{y,x_1|x_2}$ . Betaparameterschätzer sind also im Allgemeinen keine partiellen Stichprobenkorrelationen.

Allerdings gilt  $\hat{\beta}_1 = r_{y,x_1|x_2}$  genau dann, wenn  $s_y = s_{x_1}$  und zudem  $r_{y,x_1} = r_{x_1,x_2} = 0$ , 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.

Schließlich gilt obige Identität von Betaparameterschätzerkomponente und partieller Stichprobenkorrelation auch dann, wenn  $|r_{y,x_1}| = |r_{x_1,x_2}|$ , wenn also die obigen Stichprobenkorrelationen dem Betrage nach gleich sind. Dies ist in der Anwendung aber vermutlich selten der Fall.

Folgender **R**-Code demonstriert die Äquivalenz der Matrixschreibweise des Betaparameterschätzers und seiner Darstellung durch Stichprobenmittel, Stichprobenkorrelationen, Stichprobenstandardabweichungen und partielle Stichprobenkorrelationen.

```
# Daten einlesen, Modellformulierung und Modellschätzung
D = read.csv("data/13_Multiple_Regression.csv") # Daten einlesen
y = D$BDI # abhängige Variable
n = length(y) # Anzahl Datenpunkte
X = matrix(c(rep(1,n), D$Alter, D$Dauer), 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) # Tools für partielle Korrelationen
```

Loading required package: MASS

```
y12 = cbind(y,X[,2:3]) # Matrix (y, x_1, x_2)
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
```

```
Korrelationen r(y,x_1), r(y,x_2), r(x_1,x_2) : -0.87 0.64 -0.22
partielle Korrelationen r(y,x_1|x_2), r(y,x_2|x_1) : -0.97 0.92
beta_hat ALM-Schätzer : 11.44 -0.57 1.85
beta_hat partielle Korrelationen : 11.44 -0.57 1.85
```

Abbildung 13.2 visualisiert den Beispieldatensatz zusammen mit der durch den Betaparameterschätzer definierten Regressionsebene

$$f_{\hat{\beta}} : \mathbb{R}^2 \rightarrow \mathbb{R}, (x_1, x_2) \mapsto f_{\hat{\beta}}(x_1, x_2) := \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \quad (13.33)$$

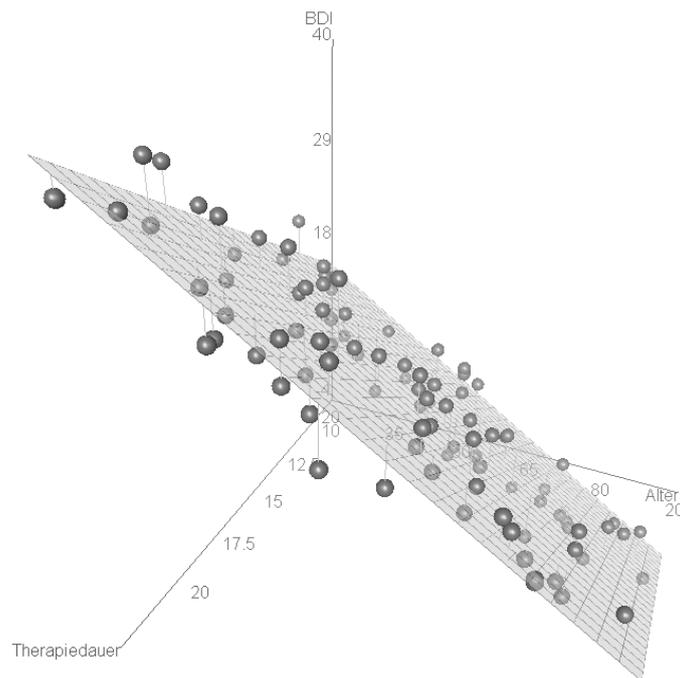


Abbildung 13.2 Visualisierung von Beispieldatensatz und Regressionebene

## 13.4 Modellevaluation

Die allgemeine Theorie der T- und F-Statistiken bietet eine Vielzahl von Möglichkeiten, verschiedenste Hypothesen im Kontext der multiplen Regression inferenzstatistisch zu evaluieren. Für das Anwendungsbeispiel könnte zum Beispiel folgende Auswahl von Kontrastgewichtsvektoren und Null- und Alternativhypothesen zunächst von Interesse sein:

$$c := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Leftrightarrow H_0 : \beta_2 = 0, H_A : \beta_2 \neq 0 \quad \text{und} \quad c := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \Leftrightarrow H_0 : \beta_3 = 0, H_A : \beta_3 \neq 0 \quad (13.34)$$

Dabei würde das Ablehnen der jeweiligen Nullhypothese jeweils inferenzstatistische Evidenz für einen Effekt des Patient:innenalters bzw. der Therapiedauer auf den BDI-Differenzwert im Kontext der Präsenz der jeweils anderen unabhängigen Variable und des Interzeptterms implizieren.

Weiterhin könnte folgender Kontrastgewichtsvektor mit folgenden Null- und Alternativhypothesen von Interesse sein:

$$c := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \Leftrightarrow H_0 : \beta_2 - \beta_3 = 0, H_A : \beta_2 - \beta_3 \neq 0 \quad (13.35)$$

In diesem Fall würde das Ablehnen der Nullhypothese inferenzstatistische Evidenz für einen differentiellen Einfluss von Patient:innenalter und Therapiedauer implizieren, je nach Vorzeichen der T-Statistik dabei einen stärkeren Effekt des Patient:innenalters oder der Therapiedauer.

Folgender **R**-Code evaluiert die angesprochenen T-Tests.

```
# Daten einlesen und Modellschätzung
D      = read.csv("_data/13_Multiple_Regression.csv") # Datensatz
y      = D$dBDI # abhängige Variable
n      = length(y) # Anzahl Datenpunkte
X      = matrix(c(rep(1,n), D$Alter, D$Dauer), 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 und 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
  tee[i] = t_num / ste[i] # T-Statistik
  pvals[i] = 2 * (1 - pt(abs(tee[i]), n-p)) # p-Wert
}

```

	Estimate	Std. Error	t value	Pr(> t )
(Interzept)	11.44	4.20	2.73	0.01
Alter	-0.57	0.04	-16.00	0.00
Dauer	1.85	0.19	9.97	0.00
Alter-Dauer	-2.42	0.18	-13.37	0.00

Der Beispieldatensatz liefert also inferenstatistische Inferenz für einen negativen Zusammenhang zwischen Therapieerfolg und Patient:innenalter und einen positiven Zusammenhang zwischen Therapieerfolg und Therapiedauer. Die Differenz der beiden Effekte ist dabei deutlich ausgeprägt.

Etwas globaler könnte man einen F-Test basierend auf einer Modellpartition mit  $p_0 := 1$  nutzen, um etwa zu klären, ob die Daten die Annahme, dass Patient:innenalter und Therapie überhaupt zur Erklärung der BDI-Differenzwertvariation beitragen, bestätigen. Folgender **R**-Code evaluiert den angesprochenen F-Test.

```
# Daten einlesen und Modellformulierung
D      = read.csv("_data/13_Multiple_Regression.csv") # Datensatz
y      = D$dBDI # abhängige Variable
n      = length(y) # Anzahl Datenpunkte
X      = matrix(c(rep(1,n), D$Alter, D$Dauer), nrow = n) # Designmatrix vollständiges Modell
p      = ncol(X) # Anzahl Parameter vollständiges Modell
p_0    = 1 # reduziertes Modell: (1_n) # Anzahl Parameter reduziertes Modell
p_1    = p - p_0 # vollständiges Modell: (1_n, x_1, x_2) # Anzahl zusätzlicher Parameter vollst. Mod.
X_0    = X[,1:p_0] # Designmatrix reduziertes Modell

# Modellschätzung und Modellevaluation
beta_hat_0 = solve(t(X_0) %*% X_0) %*% t(X_0) %*% y # Betaparameterschätzer reduziertes Modell
beta_hat    = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer vollständiges Modell
eps_hat_0   = y - X_0 %*% beta_hat_0 # Residuenvektor reduziertes Modell
eps_hat     = y - X %*% beta_hat # Residuenvektor vollständiges Modell
eh0_eh0     = t(eps_hat_0) %*% eps_hat_0 # residuelle Quadratsumme reduziertes Modell
eh_eh       = t(eps_hat) %*% eps_hat # residuelle Quadratsumme vollständiges Modell

```

```
sigsqr_hat = eh_eh/(n-p) # Varianzparameterschätzer vollständiges Modell
f           = ((eh0_eh0-eh_eh)/p_1)/sigsqr_hat # F-Statistik
pval       = 1 - pf(f,p_1,n-p) # p-Wert
```

F-statistic: 223.7 on 2 and 17 DF, p-value: 6.17950135506362e-13

Der Beispieldatensatz enthält also deutliche Evidenz dafür, dass die Aufnahme von Patient:innenalter und/oder Therapiedauer zur Erklärung der BDI-Differenzwertvariation über Patient:innen beiträgt.

Eine direkte Implementation obiger Analyse erlaubt das Zusammenspiel der **R**-Funktionen `lm()` und `summary()`, wie untenstehender **R**-Code demonstriert.

```
D = read.csv("_data/13_Multiple_Regression.csv") # Datensatz
alm = lm(dBDI ~ Alter + Dauer, data = D) # Modellformulierung, Modellschätzung
summary(alm) # Modellevaluation
```

Call:

```
lm(formula = dBDI ~ Alter + Dauer, data = D)
```

Residuals:

```
    Min       1Q   Median       3Q      Max
-4.9474 -2.0549  0.6748  1.9590  3.7926
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.43949    4.19776   2.725  0.0144 *
Alter       -0.57162    0.03573  -15.997 1.11e-11 ***
Dauer        1.85132    0.18577   9.966 1.63e-08 ***
---

```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.612 on 17 degrees of freedom

Multiple R-squared: 0.9634, Adjusted R-squared: 0.9591

F-statistic: 223.7 on 2 and 17 DF, p-value: 6.179e-13

## 13.5 Literaturhinweise

Die moderne Geschichte der multiplen Regression wird oft in den Arbeiten von Legendre (1805) und Gauss (1809) zur Berechnung von Planetenbahnen verankert. Die heutige Theorie der multiplen Regression sehr breit gefächert, so dass wir in diesem Abschnitt nur einen ersten Eindruck von ihrem Wesen vermitteln können. Weiterführende Einsichten bieten zum Beispiel Draper und Smith (1998), Hocking (2003) und Fox und Tishby (2016).

## 13.6 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 und Kovariate im Rahmen der multiplen Regression.

4. Erläutern Sie, warum  $\hat{\beta} \approx \text{Regressorkovariabilität}^{-1} \cdot \text{Regressordatenkovariabilität}$  gilt.
5. Erläutern Sie den Zusammenhang zwischen Betaparameterschätzern und Korrelationen in einem multiplen Regressionmodell mit Interzeptparameter und zwei kontinuierlichen Prädiktoren anhand der Formel

$$\hat{\beta}_1 = \frac{r_{y,x_1} - r_{y,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_y}{s_{x_1}}. \quad (13.36)$$

6. Erläutern Sie den Zusammenhang zwischen Betaparameterschätzern und partieller Korrelation in einem multiplen Regressionmodell mit Interzeptparameter und zwei kontinuierlichen Prädiktoren anhand der Formel

$$\hat{\beta}_1 = r_{y,x_1 \setminus x_2} \sqrt{\frac{1 - r_{y,x_2}^2}{1 - r_{x_1,x_2}^2}} \frac{s_y}{s_{x_1}}. \quad (13.37)$$

7.  $X \in \mathbb{R}^{n \times 2}$  sei die Designmatrix eines multiplen Regressionsmodells mit zwei Prädiktoren und Beta-parametervektor  $\beta := (\beta_1, \beta_2)^T$ . Geben Sie den Kontrastgewichtsvektor an, um die Nullhypothese  $H_0 : \beta_1 = \beta_2$  mithilfe der T-Statistik zu testen.

# 14 Kovarianzanalyse

## 14.1 Vorbemerkungen

Da mechanistische wissenschaftliche Theorien in der Regel Unterschiede zwischen Datenpunkten auf Grundlage unabhängiger Variablen (UV) vorhersagen und da diese Datenpunkte zusätzlich durch in ihnen enthaltene Zufallsfehler charakterisiert sind, unterscheiden sich die Werte einer abhängigen Variable (AV) im Allgemeinen zwischen experimentellen Einheiten. Diese Schwankungen der Werte einer AV bezeichnet man als *Datenvariabilität*. Ziel jeder Datenanalyse ist die Erklärung von Datenvariabilität durch Zerlegung in Varianzkomponenten.

Eine spezielle Art, Datenvariabilität zu quantifizieren, ist die Stichprobenvarianz, die, wie wir gesehen haben, in der Regel proportional zur total Quadratsumme ist. Das Allgemeine Lineare Modell bietet einen additiven Zugang zur Dekomposition dieser Datenvariabilität in verschiedene Komponenten. Den folgenden Überlegungen liegt datenanalytisch die in diesem Kapitel behandelte Kovarianzanalyse zugrunde.

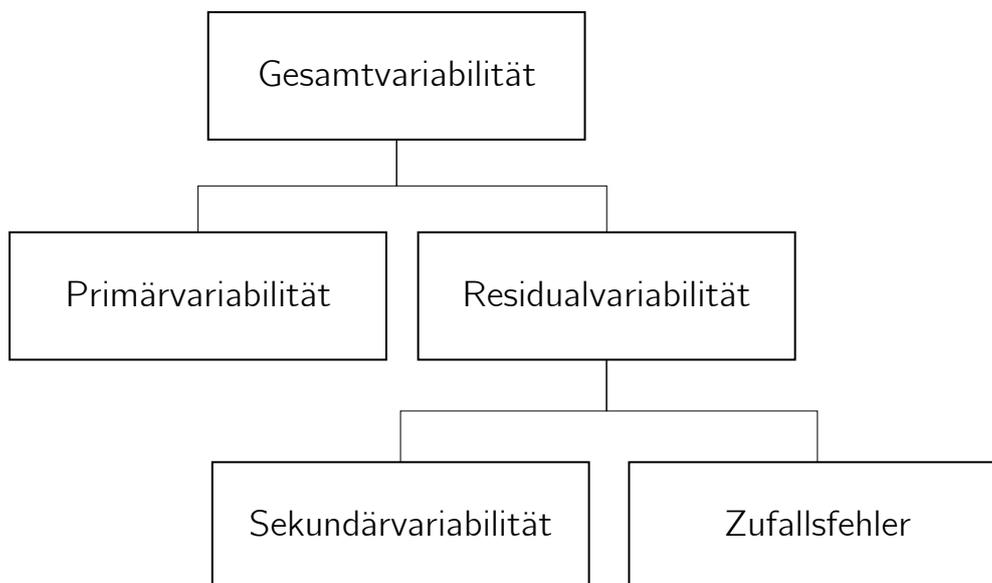


Abbildung 14.1 Variabilitätsquellen.

Einerseits lässt sich die in einem Datensatz enthaltene *Gesamtvariabilität* in *Primärvariabilität* und *Residualvariabilität* zerlegen, wobei unter Primärvariabilität systematische Veränderung der AV, die allein auf Variation der UV zurückzuführen ist, verstanden wird. Andererseits kann man die dabei übrig bleibende *Residualvariabilität* weiter in *Sekundärvariabilität* und *Zufallsfehler* zerlegen, wobei unter Sekundärvariabilität systematische

Veränderung der AV, die auf die Wirkung von unkontrollierten Störvariablen, nicht aber auf die Variation der UV zurückzuführen ist, verstanden wird. Der *Zufallsfehler* schließlich entspricht der unsystematischen Veränderung der AV, die weder auf die Variation der UV noch auf den Einfluss von Störvariablen zurückzuführen ist. Abbildung 14.1 veranschaulicht diese Begriffe.

Wie wir in Kapitel 5 gesehen haben, lassen sich (1) *faktorielle Designs*, bei denen die Designmatrix im Wesentlichen nur 1en und 0en enthält (*Einstichproben- und Zweistichproben-T-Tests* sowie *einfaktorielle und mehrfaktorielle Varianzanalyse*), und (2) *parameterische Designs*, bei denen die Designmatrix aus Spalten mit kontinuierlichen reellen Werten besteht (*einfache lineare Regression* und *multiple lineare Regression*), unterscheiden.

Mit der Kovarianzanalyse begegnen wir nun einem Beispiel für (3) *faktoriell-parametrische Designs*, bei denen die Spalten der Designmatrix sowohl faktorielle als auch parametrische Prädiktoren repräsentieren. In diesem Kontext werden die parametrischen Regressoren oft als *Kovariaten* betrachtet und der Fokus liegt auf der Beurteilung des Einflusses der faktoriellen Regressoren, also der Überprüfung möglicher Gruppenunterschiede. Wir betrachten hier exemplarisch den Fall eines Faktors und eines parametrischen Regressors.

## 14.2 Additive Kovarianzanalyse

Im Modell der additiven einfaktoriellen Kovarianzanalyse möchte man die Erwartungswertparameter  $\mu_{ij}$  der insgesamt  $n = \sum_{i=1}^I n_i$  Datenvariablen

$$y_{ij} \sim N(\mu_{ij}, \sigma^2 I_n) \quad (14.1)$$

basierend auf den Levels  $i = 1, \dots, I$  des experimentellen Faktors und den entsprechenden Werten  $x_{ij}$  für  $j = 1, \dots, n_i$  der Kovariate modellieren. Dazu wählt man für jeden Erwartungswertparameter  $\mu_{ij}$  zunächst einen Parameter  $\mu_0$ , der einen Datenvariablenunspezifischen Offset modelliert. Weiterhin wählt man einen Parameter  $\alpha_i$ , der den Beitrag des  $i$ ten Levels des experimentellen Faktors modelliert und schließlich einen Gewichtungsparemeter  $\beta_0$ , der den Beitrag der Kovariatenwerten  $x_{ij}$  zu  $\mu_{ij}$  quantifiziert.

Das Modell für den Erwartungswertparameter von  $y_{ij}$  für  $i = 1, \dots, I, j = 1, \dots, n_i$  nimmt im Fall der additiven einfaktoriellen Kovarianzanalyse also folgende Form an:

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \text{ mit } \alpha_1 := 0, \quad (14.2)$$

wobei  $\alpha_1 := 0$  wie im Falle der einfaktoriellen Varianzanalyse eine Überparameterisierung des Modells verhindert. Man beachte, dass die Form des  $ij$ ten Erwartungswertparameters damit eine einfache lineare Regression für jedes Level  $i = 1, \dots, I$  des experimentellen Faktors definiert, bei der  $\mu_0 + \alpha_i$  als Faktorlevel-spezifische Offsetparameter interpretiert werden können und der Steigungsparameter  $\beta_0$  für alle Faktorlevel identisch ist. Folgende Definition fasst diese Überlegungen formal zusammen.

**Definition 14.1** (Modell der additiven einfaktoriellen Kovarianzanalyse).  $y_{ij}$  mit  $i = 1, \dots, I, j = 1, \dots, n_i$  sei die Zufallsvariable, die den zum  $i$ ten Level des Faktors gehörenden

$j$ ten Datenpunkt modelliert und  $x_{ij}$  sei der entsprechende Wert der Kovariate. Dann hat das *Modell der additiven einfaktoriellen Kovarianzanalyse* die strukturelle Form

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, \dots, I, j = 1, \dots, n_i, \sigma^2 > 0 \quad (14.3)$$

und die Datenverteilungsform

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. f\"ur } i = 1, \dots, I, j = 1, \dots, n_i, \quad (14.4)$$

wobei

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \text{ mit } \alpha_1 := 0. \quad (14.5)$$

•

Aus der strukturellen bzw. Datenverteilungsform dieses Modells lässt sich ebenfalls seine Designmatrixform ableiten, die in folgendem Theorem angegeben ist. Wir verzichten hierbei auf einen Beweis, da es sich direkt mit den Regeln der Matrixmultiplikation ergibt.

**Theorem 14.1** (Designmatrixform der additiven einfaktoriellen Kovarianzanalyse). *Gegeben sei die strukturelle Form eines additiven einfaktoriellen Kovarianzanalysemodells mit Anzahl der Faktorlevel  $I$  und es sei  $n := \sum_{i=1}^I n_i$  die Gesamtanzahl an Datenvariablen. Dann hat dieses Modell die Designmatrixform*

$$y \sim N(X\beta, \sigma^2 I_n) \quad (14.6)$$

mit

$$y := \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & 0 & x_{11} \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & x_{1n_1} \\ 1 & 1 & 0 & x_{21} \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & 0 & x_{2n_2} \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 1 & x_{I1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 1 & x_{In_2} \end{pmatrix} \in \mathbb{R}^{n \times (I+1)}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_I \\ \beta_0 \end{pmatrix} \in \mathbb{R}^{I+1} \quad (14.7)$$

und  $\sigma^2 > 0$ .

◦

## Anwendungsbeispiel

Unser Anwendungsbeispiel benutzt als experimentellen Faktor zwei Studiengruppen ( $I := 2$ ,  $i = 1, 2$ ) mit Face-to-Face- vs. Online-Psychotherapie, als parametrischen Regressor bzw. Kovariate ( $x_{ij}$ ,  $j = 1, \dots, n_i$ ) die Dauer der Depressionssymptomatik zu Beginn der Intervention und als abhängige Variable bzw. primäres Ergebnismaß ( $y_{ij}$ ,  $j = 1, \dots, n_i$ ) BDI-Differenzwerte im Sinne einer Vorher-Nachher-Messung im Rahmen der Interventionsstudie.

Tabelle 14.1 zeigt einen Beispieldatensatz, bestehend aus Datenwerten von  $n = 32$  Patient:innen, wobei “Therapie” die Therapieform, “Dauer” die Dauer der Depressionssymptomatik und “dBDI” die BDI-Differenzwerte in der Pre-Post-Interventionsstudie bezeichnen. Abbildung 14.2 stellt diesen Datensatz graphisch dar.

**Tabelle 14.1** BDI-Differenzwerte in Abhängigkeit von Therapiegruppe und Dauer der Depressionssymptomatik (jeweils erste 10 Patient:innen für beide Gruppen).

	Therapie	Dauer	dBDI
1	F2F	3.7	6
2	F2F	5.4	6
3	F2F	3.3	7
4	F2F	8.2	7
5	F2F	5.7	4
6	F2F	3.4	8
7	F2F	6.0	2
8	F2F	6.5	3
9	F2F	6.2	8
10	F2F	4.4	8
17	ONL	5.0	12
18	ONL	6.9	9
19	ONL	6.6	7
20	ONL	6.2	7
21	ONL	6.8	9
22	ONL	6.6	10
23	ONL	5.1	9
24	ONL	1.0	14
25	ONL	6.2	10
26	ONL	4.9	12

Zum Zwecke der Datenanalyse betrachten wir nun ein Modell ohne Berücksichtigung der Kovariate (Modell 1)

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \text{ mit } \mu_{ij} := \mu_0 + \alpha_i \text{ und } \alpha_1 := 0 \text{ und } \sigma^2 > 0 \quad (14.8)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \end{pmatrix} \text{ und } \sigma^2 > 0 \quad (14.9)$$

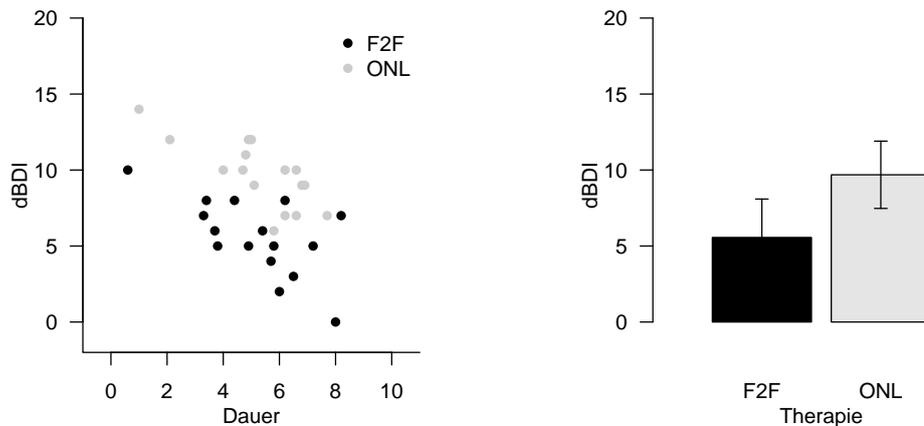


Abbildung 14.2 Beispieldatensatz zur additiven Kovarianzanalyse.

sowie eines mit additiver Berücksichtigung der Kovariate (Modell 2)

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \text{ mit } \mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \text{ und } \alpha_1 := 0 \text{ und } \sigma^2 > 0 \quad (14.10)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & 0 & x_{11} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{1n_1} \\ 1 & 1 & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n_2} \end{pmatrix}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_0 \end{pmatrix} \text{ und } \sigma^2 > 0. \quad (14.11)$$

Wie man anhand der Designmatrizen erkennt, handelt es sich hierbei um ein Modell der einfaktoriellen Varianzanalyse (Modell 1) und ein Modell der additiven Kovarianzanalyse (Modell 2). Folgender **R**-Code implementiert diese beiden Modelle im Sinne von Modellformulierung, Parameterschätzung und schließlich Modellevaluation in Form eines Tests für den Effekt des kategorischen Faktors Therapieform (Face-to-Face- vs. Online-Psychotherapie) unter Berücksichtigung der Kovariate (Dauer der Depressionssymptomatik).

```
fname = "_data/14_Kovarianzanalyse-1.csv" # Datensatzdateiname
D = read.table(fname, sep = ",", header = TRUE) # Datensatz
n_i = c(sum(D$Therapie == "F2F"), sum(D$Therapie == "ONL")) # Anzahl Datenpunkte pro Gruppe
y = D$dBDI # Datenvektor
n = length(y) # Gesamtanzahl Datenpunkte
XS = list() # Modell 1 und 2 Liste
CS = list() # Kontrastgewichtsvektorenliste
XS[[1]] = matrix(c(rep(1, n_i[1]), rep(1, n_i[2]), # \mu_0 Regressor
                  rep(0, n_i[1]), rep(1, n_i[2])), nrow = n) # \alpha_2 Regressor
XS[[2]] = matrix(c(rep(1, n_i[1]), rep(1, n_i[2]), # \mu_0 Regressor
                  rep(0, n_i[1]), rep(1, n_i[2]), # \alpha_2 Regressor
                  D$Dauer), nrow = n) # x Regressor
CS[[1]] = matrix(c(0, 1), nrow = 2) # Kontrastgewichtsvektor \alpha_2
CS[[2]] = matrix(c(0, 1, 0), nrow = 3) # Kontrastgewichtsvektor \alpha_2
B = list() # Betaparameterschätzerliste
S = rep(NA, 2) # Varianzparameterschätzervektor
T = rep(NA, 2) # T-Statistik-Vektor
for(i in 1:2){
  X = XS[[i]] # Designmatrix
  p = ncol(X) # Anzahl Betaparameter
  beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
  eps_hat = y - X %*% beta_hat # Residuenvektor
  sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
  c = CS[[i]] # Kontrastgewichtsvektor
  t_num = t(c) %*% beta_hat # Zähler der Zweistichproben-T-Teststatistik
  t_den = sqrt(sigsq_hat %*% t(c) %*% solve(t(X) %*% X) %*% c) # Nenner der Zweistichproben-T-Teststatistik
}
```

```

t      = t_num/t_den      # Wert der Zweistichproben-T-Teststatistik
B[[i]] = beta_hat        # Betaparameterschätzer
S[i]   = sigsq_hat       # Varianzparameterschätzer
T[i]   = t                # T-Statistik
}

```

```

Modell 1 | Betaparameterschätzer      :  5.56  4.12
Modell 2 | Betaparameterschätzer      : 10.05  4.2 -0.86
Modell 1 | Varianzparameterschätzer   :  5.65
Modell 2 | Varianzparameterschätzer   :  3.14
Modell 1 | T-Statistik ONL-Therapie    :  4.91
Modell 2 | T-Statistik ONL-Therapie    :  6.7

```

In Modell 2 ist der Varianzparameterschätzer kleiner als in Modell 1. Dies weist darauf hin, dass der parametrische Regressor tatsächlich Datenvarianz erklärt. Die Betaparameterschätzer für den ONL-Therapieeffekt sind in Modell 1 und Modell 2 ähnlich. Zusammen hat dies zur Folge, dass die T-Statistik für den ONL-Therapieeffekt im Sinne eines Signal-zu-Rauschen-Verhältnis in Modell 2 höher ist.

Prinzipiell kann man die beobachteten Daten für den geschätzten Effekt der Dauer durch

$$\hat{y}_{ij}^{\text{adj}} = y_{ij} - \hat{\beta}_0 x_{ij} \quad (14.12)$$

“korrigieren” und erhält so einen “adjustierten Datensatz” (siehe Abbildung 14.3).

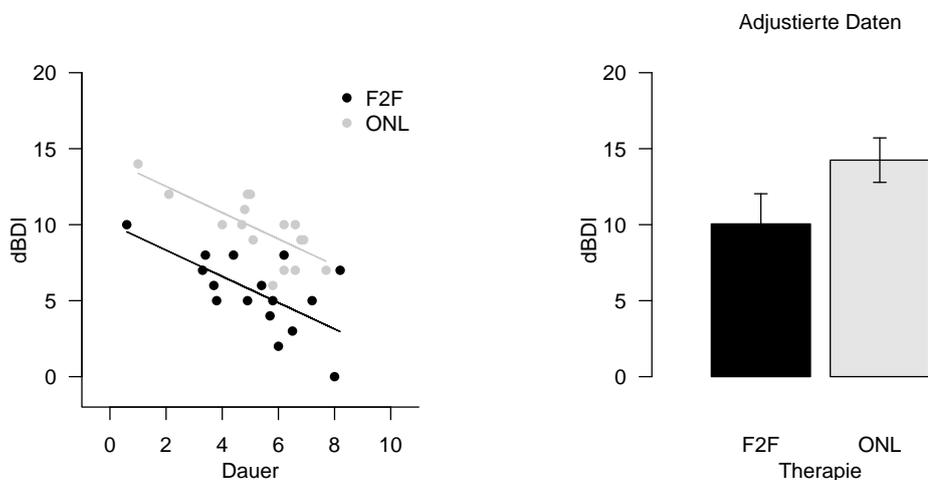


Abbildung 14.3 Ergebnisse der additiven Kovarianzanalyse.

## 14.3 Kovarianzanalyse mit Interaktion

Wie im Modell der additiven einfaktoriellen Kovarianzanalyse möchte man im Modell der einfaktoriellen Kovarianzanalyse mit Interaktion die Erwartungswertparameter  $\mu_{ij}$  der insgesamt  $n = \sum_{i=1}^I n_i$  Datenvariablen

$$y_{ij} \sim N(\mu_{ij}, \sigma^2 I_n) \quad (14.13)$$

basierend auf den Leveln  $i = 1, \dots, I$  des experimentellen Faktors und den entsprechenden Werten  $x_{ij}$  für  $j = 1, \dots, n_i$  der Kovariate modellieren. Über das Szenario der additiven einfaktoriellen Kovarianzanalyse hinaus möchte man dabei explizit einen Faktorlevel-spezifischen Einfluss der Kovariate auf die Erwartungswertparameter zulassen.

Dazu wählt man für jeden Erwartungswertparameter  $\mu_{ij}$  zunächst wieder einen Parameter  $\mu_0$ , der einen Datenvariablen-unspezifischen Offset modelliert. Weiterhin wählt man einen Parameter  $\alpha_i$ , der den Beitrag des  $i$ ten Levels des experimentellen Faktors modelliert. Den Wichtungsparemetr des Beitrags der Kovariaten  $x_{ij}$  zu  $\mu_{ij}$  modelliert man nun hier mithilfe eines Faktorlevel-unspezifischen Parameter  $\beta_0$  und eines Faktorlevel-spezifischen Parameters  $\gamma_i$  für  $i = 1, \dots, I$ . Das Modell für den Erwartungswertparameter von  $y_{ij}$  für  $i = 1, \dots, I, j = 1, \dots, n_i$  nimmt im Fall der einfaktoriellen Kovarianzanalyse mit Interaktion damit folgende Form an:

$$\mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \text{ mit } \alpha_1 := \gamma_1 := 0, \quad (14.14)$$

wobei  $\alpha_1 := \gamma_1 := 0$  wie im Fall der zweifaktoriellen Varianzanalyse mit Interaktion eine Überparametrisierung des Modells verhindert. Man beachte, dass die Form des  $ij$ ten Erwartungswertparameters damit eine einfache lineare Regression für jedes Level  $i = 1, \dots, I$  des experimentellen Faktors definiert, wobei  $\mu_0 + \alpha_i$  als Faktorlevel-spezifische Offsetparameter interpretiert werden können und sich die Steigungsparameter  $\beta_0 + \gamma_i$  über die Faktorlevel unterscheiden können. Es werden also Unterschiede (zwischen den Faktorleveln) von Unterschieden (zwischen den Erwartungswertparametern eines Faktorlevels in Abhängigkeit vom Wert der Kovariate), d.h. Interaktionen, mit modelliert. Folgende Definition fasst diese Überlegungen erneut zusammen.

**Definition 14.2** (Modell der einfaktoriellen Kovarianzanalyse mit Interaktion).  $y_{ij}$  mit  $i = 1, \dots, I, j = 1, \dots, n_i$  sei die Zufallsvariable, die den zum  $i$ ten Level des Faktors gehörenden  $j$ ten Datenpunkt modelliert und  $x_{ij}$  sei der entsprechende Wert der Kovariate. Dann hat das *Modell einfaktoriellen Kovarianzanalyse mit Interaktion* die strukturelle Form

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, I, j = 1, \dots, n_i, \sigma^2 > 0 \quad (14.15)$$

und die Datenverteilungsform

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. für } i = 1, \dots, I, j = 1, \dots, n_i, \quad (14.16)$$

wobei

$$\mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \text{ mit } \alpha_1 := \gamma_1 := 0. \quad (14.17)$$

•

Aus der strukturellen bzw. Datenverteilungsform lässt sich wiederum die Designmatrixform des Modells ableiten, die in folgendem Theorem angegeben ist. Wir verzichten auch hier auf einen Beweis, da es sich direkt mit den Regeln der Matrixmultiplikation ergibt.

**Theorem 14.2** (Designmatrixform der einfaktoriellen Kovarianzanalyse mit Interaktion). Gegeben sei die strukturelle Form eines einfaktoriellen Kovarianzanalysemodells mit Interaktion mit Anzahl der Faktorlevel  $I$  und es sei  $n := \sum_{i=1}^I n_i$  die Gesamtanzahl an Datenvariablen. Dann hat dieses Modell die Designmatrixform

$$y \sim N(X\beta, \sigma^2 I_n) \tag{14.18}$$

mit

$$y := \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & 0 & x_{11} & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & x_{1n_1} & 0 & \dots & 0 \\ 1 & 1 & 0 & x_{21} & x_{21} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & x_{2n_2} & x_{2n_2} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & x_{I1} & 0 & \dots & x_{I1} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & x_{In_2} & 0 & \dots & x_{In_2} \end{pmatrix} \in \mathbb{R}^{n \times (2I)}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_I \\ \beta_0 \\ \gamma_2 \\ \vdots \\ \gamma_I \end{pmatrix} \in \mathbb{R}^{2I} \tag{14.19}$$

und  $\sigma^2 > 0$ .

◦

### Anwendungsbeispiel

Unser Anwendungsbeispiel benutzt als experimentellen Faktor erneut zwei Studiengruppen ( $I := 2, i = 1, 2$ ) mit Face-to-Face- vs. Online-Psychotherapie, als parametrischen Regressor bzw. Kovariate ( $x_{ij}, j = 1, \dots, n_i$ ) hier nun die Digitalaffinität der Studienteilnehmer und als abhängige Variable bzw. primäres Ergebnismaß ( $y_{ij}, j = 1, \dots, n_i$ ) wie zuvor Pre/Post-BDI-Differenzwerte.

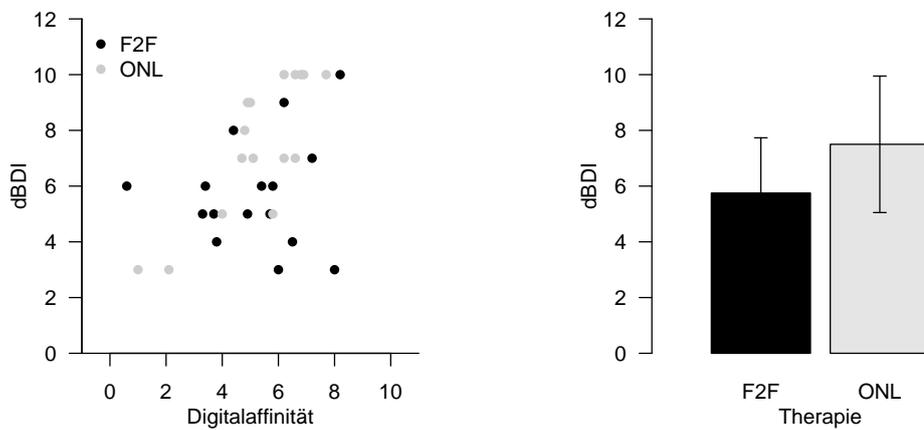
Tabelle 14.2 zeigt einen Beispieldatensatz, bestehend aus Datenwerten von  $n = 32$  Patient:innen, wobei "Therapie" die Therapieform, "Dig.Affin" die Digitalaffinität und "dBDI" die BDI-Differenzwerte in der Pre-Post-Interventionsstudie bezeichnen. Abbildung 14.4 stellt diesen Datensatz graphisch dar.

**Tabelle 14.2** BDI-Differenzwerte in Abhängigkeit von Therapiegruppe und Digitalaffinität (jeweils erste 10 Patient:innen für beide Gruppen).

	Therapie	Dig.Affin	dBDI
1	F2F	3.7	5
2	F2F	5.4	6
3	F2F	3.3	5
4	F2F	8.2	10
5	F2F	5.7	5
6	F2F	3.4	6
7	F2F	6.0	3
8	F2F	6.5	4
9	F2F	6.2	9

**Tabelle 14.2** BDI-Differenzwerte in Abhängigkeit von Therapiegruppe und Digitalaffinität (jeweils erste 10 Patient:innen für beide Gruppen).

	Therapie	Dig.Affin	dBDI
10	F2F	4.4	8
17	ONL	5.0	9
18	ONL	6.9	10
19	ONL	6.6	7
20	ONL	6.2	7
21	ONL	6.8	10
22	ONL	6.6	10
23	ONL	5.1	7
24	ONL	1.0	3
25	ONL	6.2	10
26	ONL	4.9	9



**Abbildung 14.4** Beispieldatensatz zur Kovarianzanalyse mit Interaktion.

Zum Zwecke der Datenanalyse betrachten wir nun ein Modell mit additiver Berücksichtigung der Kovariate (Modell 1)

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \text{ mit } \mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \text{ und } \alpha_1 := 0 \text{ und } \sigma^2 > 0 \quad (14.20)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & 0 & x_{11} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{1n_1} \\ 1 & 1 & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n_2} \end{pmatrix}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_0 \end{pmatrix} \text{ und } \sigma^2 > 0 \quad (14.21)$$

sowie eines mit additiver und interaktiver Berücksichtigung der Kovariate (Modell 2)

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \text{ mit } \mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \text{ und } \alpha_1 := \gamma_1 := 0 \text{ und } \sigma^2 > 0$$

(14.22)

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & 0 & x_{11} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_{1n_1} & 0 \\ 1 & 1 & x_{21} & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n_2} & x_{2n_2} \end{pmatrix}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_0 \\ \gamma_2 \end{pmatrix} \text{ und } \sigma^2 > 0 \quad (14.23)$$

Wie man anhand der Designmatrizen erkennt, handelt es sich hierbei um ein Modell der additiven Kovarianzanalyse (Modell 1) und ein Modell der einfaktoriellen Kovarianzanalyse mit Interaktion (Modell 2). Folgender **R**-Code implementiert diese beiden Modelle im Sinne von Modellformulierung, Parameterschätzung und schließlich Modellevaluation in Form eines Zweistichproben-T-Tests für den Effekt des kategorischen Faktors Therapieform (Face-to-Face- vs. Online-Psychotherapie) unter Berücksichtigung der Kovariate (Digitalaffinität).

```

1060 fname = file.path("./Daten/Kovarianzanalyse-2.csv") # Datensatzdateiname
1061 D = read.table(fname, sep = ",", header = TRUE) # Datensatz
1062 n_i = c(sum(D$Therapie == "F2F"), sum(D$Therapie == "ONL")) # Anzahl Datenpunkte pro Gruppe
1063 y = D$dbDI # Daten
1064 n = length(y) # Gesamtanzahl Datenpunkte
1065 XS = list() # Model 1 und 2 Liste
1066 CS = list() # Kontrastgewichtsvektorenliste
1067 XS[[1]] = matrix(c(rep(1,n_i[1]), rep(1,n_i[2]), # \mu_0 Regressor
1068 rep(0,n_i[1]), rep(1,n_i[2]), # \alpha_2 Regressor
1069 D$Dig.Affin, nrow = n) # \beta_2 Regressor
1070 XS[[2]] = matrix(c(rep(1,n_i[1]), rep(1,n_i[2]), # \mu_0 Regressor
1071 rep(0,n_i[1]), rep(1,n_i[2]), # \alpha_2 Regressor
1072 D$Dig.Affin, # \beta_2 Regressor
1073 D$Dig.Affin*(c(rep(0,n_i[1]),rep(1,n_i[2]))), # \gamma_2 Regressor
1074 nrow = n) # Anzahl Zeilen der Designmatrix
1075 CS[[1]] = matrix(c(0,1,0), nrow = 3) # Kontrastgewichtsvektor \alpha_2
1076 CS[[2]] = matrix(c(0,1,0,0), nrow = 4) # Kontrastgewichtsvektor \alpha_2
1077 B = list() # Betaparameterschätzerliste
1078 S = rep(NaN,2) # Varianzparameterschätzervektor
1079 T = rep(NaN,2) # T-Statistik-Vektor
1080 for(i in 1:2){ # Iteration über Modelle
1081 X = XS[[i]] # Designmatrix
1082 p = ncol(X) # Anzahl Betaparameter
1083 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
1084 eps_hat = y - X %*% beta_hat # Residuenvektor
1085 sigsq_hat = (t(eps_hat) %*% eps_hat)/(n-p) # Varianzparameterschätzer
1086 c = CS[[i]] # Kontrastgewichtsvektor
1087 t_num = t(c) %*% beta_hat # Zähler der Zweistichproben-T-Teststatistik
1088 t_den = sqrt(sigsq_hat %*% t(c) %*% solve(t(X) %*% X) %*% c) # Nenner der Zweistichproben-T-Teststatistik
1089 t = t_num/t_den # Wert der Zweistichproben-T-Teststatistik
1090 B[[i]] = beta_hat # Betaparameterschätzer
1091 S[i] = sigsq_hat # Varianzparameterschätzer
1092 T[i] = t # T-Statistik
1093 }

```

Abbildung 14.5

Modell 1   Betaparameterschätzer	:	2.73	1.7	0.58	
Modell 2   Betaparameterschätzer	:	5.03	-3.47	0.14	0.99
Modell 1   Varianzparameterschätzer	:	3.92			

Modell 2 | Varianzparameterschätzer : 3.15  
 Modell 1 | T-Statistik ONL-Therapie : 2.43  
 Modell 2 | T-Statistik ONL-Therapie : -1.79

In Modell 1 wird der ONL-Therapieeffekt positiv, in Modell 2 negativ geschätzt. Entsprechend ist die T-Statistik für ONL-Therapie in Modell 1 positiv, in Modell 2 negativ. Die wahren, aber unbekanntem Betaparameterwerte sind  $\beta := (5, -3, 0, 1)^T$ , d.h. die Abhängigkeit von Digitalaffinität kommt bei ONL zum Tragen ( $\beta_4 = 1$ ), bei F2F hingegen nicht ( $\beta_3 = 0$ ). Vor diesem Hintergrund ist der Haupteffekt der Therapieart schwierig zu beurteilen: ONL-Therapie ist bei niedriger Digitalaffinität nicht so wirksam wie F2F-Therapie, bei hoher Digitalaffinität aber wirksamer als F2F-Therapie (siehe Abbildung 14.6).

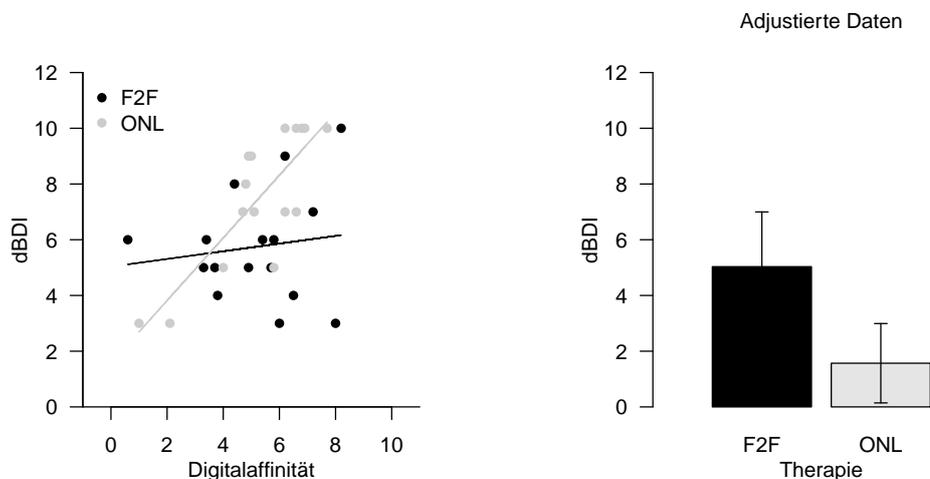


Abbildung 14.6 Ergebnisse der Kovarianzanalyse mit Interaktion.

## 14.4 Selbstkontrollfragen

1. Erläutern Sie die Begriffe der Gesamt-/Primär-/Sekundär- und Residualvariabilität.
2. Geben Sie jeweils zwei Beispiele für parametrische bzw. faktorielle ALM-Designs an.
3. Geben Sie die Definition des Modells der additiven einfaktoriellen Kovarianzanalyse (EKA) wieder.
4. Erläutern Sie die Bedeutung der Parameter  $\mu_0$ ,  $\alpha_i$  und  $\beta_0$  im Modell der additiven EKA.
5. Geben Sie die Designmatrixform des Modells der additiven EKA für  $I := 2$  und  $n_1 = n_2 = 3$  an.
6. Erläutern Sie den Unterschied zwischen dem Modell der einfaktoriellen Varianzanalyse und dem Modell der additiven einfaktoriellen Kovarianzanalyse.
7. Geben Sie die Definition des Modells der einfaktoriellen Kovarianzanalyse (EKA) mit Interaktion wieder.
8. Erläutern Sie die Bedeutung der Parameter  $\mu_0$ ,  $\alpha_i$ ,  $\beta_0$  und  $\gamma_i$  im Modell der EKA mit Interaktion.
9. Geben Sie die Designmatrixform des Modells der EKA mit Interaktion für  $I := 2$  und  $n_1 = n_2 = 3$  an.
10. Erläutern Sie den Unterschied zwischen dem Modell der additiven einfaktoriellen Kovarianzanalyse und dem Modell der einfaktoriellen Kovarianzanalyse mit Interaktion.

# Referenzen

- Aitken, A. C. 1936. „IV.—On Least Squares and Linear Combination of Observations“. *Proceedings of the Royal Society of Edinburgh* 55: 42–48. <https://doi.org/10.1017/S0370164600014346>.
- Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience.
- Baba, Kunihiro, Ritei Shibata, und Masaaki Sibuya. 2004. „Partial Correlation and Conditional Correlation as Measures of Conditional Independence“. *Australian <Html\_ent Glyph="&" Ascii="&"/> New Zealand Journal of Statistics* 46 (4): 657–64. <https://doi.org/10.1111/j.1467-842X.2004.00360.x>.
- Box, Joan Fisher. 1981. „Gosset, Fisher, and the t Distribution“. *The American Statistician* 35 (2): 61. <https://doi.org/10.2307/2683142>.
- Bravais, A. 1844. *Analyse Mathématique : Sur Les Probabilités Des Erreurs de Situation d'un Point*. Paris.
- Caley, Arthur. 1858. „A Memoir on the Theory of Matrices“. *Philosophical Transactions of the Royal Society of London* 148 (Dezember): 17–37. <https://doi.org/10.1098/rstl.1858.0002>.
- Christensen, Ronald. 2011. *Plane Answers to Complex Questions*. Springer Texts in Statistics. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4419-9816-3>.
- Cohen, Jacob. 1968. „Multiple Regression as a General Data-Analytic System.“ *Psychological Bulletin* 70 (6, Pt.1): 426–43. <https://doi.org/10.1037/h0026714>.
- DeGroot, Morris H., und Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Draper, Norman, und Harry Smith. 1998. *Applied Regression Analysis*. Wiley-Interscience.
- Edgeworth, F. Y. 1892. „The Law of Error and Correlated Averages“. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 34 (210): 429–38. <https://doi.org/10.1080/14786449208620355>.
- Everitt, Brian, und David C. Howell, Hrsg. 2005. *Encyclopedia of Statistics in Behavioral Science*. Hoboken, N.J: John Wiley & Sons.
- Fisher, R. A. 1924. „The Distribution of the Partial Correlation Coefficient.“ *Metron* 3: 329–32.
- . 1925a. „Applications of "Student's" Distribution“. *Metron* 5: 90–104.
- . 1925b. *Statistical Methods for Research Workers*. Oliver & Boyd.
- . 1925c. „Theory of Statistical Estimation“. *Mathematical Proceedings of the Cambridge Philosophical Society* 22 (5): 700–725. <https://doi.org/10.1017/S0305004100009580>.
- . 1935. *The Design of Experiments*. 1. ed. New York: Hafner Press.
- Foulley, JL. 1993. „A Simple Argument Showing How to Derive Restricted Maximum Likelihood“.
- Fox, Roy, und Naftali Tishby. 2016. „Minimum-Information LQG Control Part I: Memoryless Controllers“. In, 5610–16. IEEE. <https://doi.org/10.1109/CDC.2016.7799131>.

- Galton, Francis. 1886. „Regression Towards Mediocrity in Hereditary Stature.“ *The Journal of the Anthropological Institute of Great Britain and Ireland* 15: 246. <https://doi.org/10.2307/2841583>.
- . 1890. „Kinship and Correlation“. *Statistical Science* 4 (2). <https://doi.org/10.1214/ss/1177012581>.
- Gauss, Carl Friedrich. 1809. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Cambridge: Cambridge University Press.
- Georgii, Hans-Otto. 2009. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 4., überarb. und erw. Aufl. De-Gruyter-Lehrbuch. Berlin: de Gruyter.
- Harville, David A. 1977. „Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems“. *Journal of the American Statistical Association* 72 (358): 320. <https://doi.org/10.2307/2286796>.
- Hocking, Ronald. 2003. *Methods and Applications of Linear Models - Regression and the Analysis of Variance*. Wiley.
- Imbens, Guido, und 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): 146–49. <https://doi.org/10.1080/00031305.1976.10479163>.
- Legendre, A. M. 1805. *Nouvelles Methodes Pour La Determination Des Orbites Des Cometes*. Didot Paris.
- Lehmann, Erich L. 2011. *Fisher, Neyman, and the Creation of Classical Statistics*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4419-9500-1>.
- Lindholm, Mathias, und Felix Wahl. 2020. „On the Variance Parameter Estimator in General Linear Models“. *Metrika* 83 (2): 243–54. <https://doi.org/10.1007/s00184-019-00751-4>.
- Mardia, K. V., J. T. Kent, und J. M. Bibby. 1979. *Multivariate Analysis*. Probability and Mathematical Statistics. London ; New York: Academic Press.
- Neyman, J. 1935. „On the Problem of Confidence Intervals“. *The Annals of Mathematical Statistics* 6 (3): 111–16. <https://doi.org/10.1214/aoms/1177732585>.
- Neyman, Jerzy. 1937. „Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability“. *Statistical Stimation*.
- Neyman, J., und E. S. Pearson. 1928. „On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I“. *Biometrika* 20A (1/2): 175. <https://doi.org/10.2307/2331945>.
- Patterson, H. D., und R. Thompson. 1971. „Recovery of Inter-Block Information When Block Sizes Are Unequal“. *Biometrika* 58 (3): 545–54. <https://doi.org/10.1093/biomet/58.3.545>.
- Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K. ; New York: Cambridge University Press.
- Pearson, Karl. 1895. „Note on Regression and Inheritance in the Case of Two Parents“. *Proceedings of the Royal Society of London* 5: 240–42. <https://www.jstor.org/stable/115794>.
- . 1896. „Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia“. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 18: 253–318. <https://www.jstor.org/stable/90707>.
- . 1900. „On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling“. *The London, Edinburgh, and*

- Dublin Philosophical Magazine and Journal of Science* 50 (302): 157–75. <https://doi.org/10.1080/14786440009463897>.
- . 1901. „On Lines and Planes of Closest Fit to Systems of Points in Space“. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559–72. <https://doi.org/10.1080/14786440109462720>.
- . 1920. „Notes on the History of Correlation“. *Biometrika* 13 (1): 25–45. <https://doi.org/10.1093/biomet/13.1.25>.
- Plackett, R. L. 1949. „A Historical Note on the Method of Least Squares“. *Biometrika* 36 (3/4): 458. <https://doi.org/10.2307/2332682>.
- Rencher, Alvin C., und G. Bruce Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. Hoboken, N.J: Wiley-Interscience.
- Scheffé, Henry. 1959. *The Analysis of Variance*. Wiley classics library ed. A Wiley Publication in Mathematical Statistics. New York: Wiley-Interscience Publication.
- Seal, Hilary L. 1967. „Studies in the History of Probability and Statistics. XV: The Historical Development of the Gauss Linear Model“. *Biometrika* 54 (1/2): 1. <https://doi.org/10.2307/2333849>.
- Searle, S. R. 1971. *Linear Models*. Wiley Classics Library. New York, NY: Wiley.
- Searle, S. R., und Marvin H. J. Gruber. 2017. *Linear Models*. Second edition. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley.
- Searle, Shayle. 1982. *Matrix Algebra Useful for Statistics*. Wiley-Interscience.
- Seber, G. A. F., und A. J. Lee. 2003. *Linear Regression Analysis*. 2nd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience.
- Starke, Ludger, und Dirk Ostwald. 2017. „Variational Bayesian Parameter Estimation Techniques for the General Linear Model“. *Frontiers in Neuroscience* 11 (September). <https://doi.org/10.3389/fnins.2017.00504>.
- Stigler, Stephen M. 1981. „Gauss and the Invention of Least Squares“. *The Annals of Statistics* 9 (3). <https://doi.org/10.1214/aos/1176345451>.
- . 1986. *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, Mass: Belknap Press of Harvard University Press.
- Strang, Gilbert. 2009. *Introduction to Linear Algebra*. Cambridge University Press.
- Student. 1908. „The Probable Error of a Mean“. *Biometrika* 6 (1): 1–25.
- Tong, Y. L. 1990. *Multivariate Normal Distribution*. New York: Springer.
- Verbyla, A. P. 1990. „A Conditional Derivation of Residual Maximum Likelihood“. *Australian Journal of Statistics* 32 (2): 227–30. <https://doi.org/10.1111/j.1467-842X.1990.tb01015.x>.
- Wilks, S. S. 1938. „The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses“. *The Annals of Mathematical Statistics* 9 (1): 60–62. <https://doi.org/10.1214/aoms/1177732360>.
- Wright, Sewall. 1921. „Correlation and Causation“. *Journal of Agriculture Research* 20 (7): 557–85.
- Yule, G. Udny. 1907. „On the Theory of Correlation for Any Number of Variables, Treated by a New System of Notation“. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 79 (529): 182–93. <https://www.jstor.org/stable/92723>.
- 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>.