

Das Allgemeine Lineare Modell

Dirk Ostwald

2023-07-09

Inhaltsverzeichnis

Vorwort	3
1 Regression	4
1.1 Methode der kleinsten Quadrate	5
1.2 Einfache lineare Regression	10
1.3 Literaturhinweise	15
1.4 Selbstkontrollfragen	15
2 Korrelation	16
2.1 Grundlagen	16
2.2 Korrelation und Bestimmtheitsmaß	20
2.3 Korrelation und linear-affine Abhängigkeit	24
2.4 Literaturhinweise	26
2.5 Selbstkontrollfragen	27
3 Matrizen	28
3.1 Definition	28
3.2 Grundlegende Matrixoperationen	30
3.3 Matrixmultiplikation	35
3.4 Matrixinversion	40
3.5 Determinanten	43
3.6 Spezielle Matrizen	47
3.7 Literaturhinweise	50
3.8 Selbstkontrollfragen	50
4 Normalverteilungen	52
4.1 Konstruktion	52
4.2 Definition	55
4.3 Transformationen	56
4.4 Sphärizität	58
4.5 Marginale und bedingte Verteilungen	59
4.6 Literaturhinweise	62
4.7 Selbstkontrollfragen	62
5 Modellformulierung	63
5.1 Allgemeine Theorie	63
5.2 Identifizierbarkeit und Schätzbarkeit	66
5.3 Designspektrum	69
5.4 Literaturhinweise	70
5.5 Selbstkontrollfragen	70
6 Parameterschätzung	71
6.1 Betaparameterschätzung	71

6.2	Varianzparameterschätzung	73
6.3	Unabhängig identisch normalverteilte Zufallsvariablen	74
6.4	Einfache lineare Regression	75
6.5	Frequentistische Schätzerverteilungen	78
6.6	Literaturhinweise	80
6.7	Selbstkontrollfragen	80
7	T-Statistiken	82
7.1	Definition und Beispiele	82
7.2	Konfidenzintervalle für Betaparameterkomponenten	86
7.3	Literaturhinweise	88
7.4	Selbstkontrollfragen	88
8	F-Statistiken	90
8.1	Likelihood-Quotienten-Statistiken	90
8.2	Definition und Verteilung	92
8.3	Literaturhinweise	96
8.4	Selbstkontrollfragen	96
9	T-Tests	98
9.1	Einstichproben-T-Tests	98
9.2	Zweistichproben-T-Tests	102
9.3	Literaturhinweise	107
9.4	Selbstkontrollfragen	107
10	Einfaktorielle Varianzanalyse	109
10.1	Anwendungsszenario	109
10.2	Modellformulierung	110
10.3	Modellschätzung	114
10.4	Modellevaluation	117
10.5	Literaturhinweise	127
10.6	Selbstkontrollfragen	127
11	Zweifaktorielle Varianzanalyse	129
11.1	Anwendungsszenario	129
11.2	Modellformulierung	133
11.3	Modellschätzung	141
11.4	Modellevaluation	146
11.5	Literaturhinweise	151
11.6	Selbstkontrollfragen	151
12	Partielle Korrelation	152
12.1	Motivation	152
12.2	Bedingte Korelation	154
12.3	Partielle Korrelation	157
12.4	Literaturhinweise	160
12.5	Selbstkontrollfragen	160
13	Multiple Regression	161
13.1	Anwendungsszenario	161
13.2	Modellformulierung	162

13.3	Modellschätzung	163
13.4	Modellevaluation	170
13.5	Literaturhinweise	172
13.6	Selbstkontrollfragen	172
	Referenzen	174

Vorwort

Das Allgemeine Lineare Modell ist ein integrativer Zugang zu einer Reihe typischer inferenzstatistischer Verfahren wie einfacher und multipler linearer Regression, T-Tests, Varianzanalysen und Kovarianzanalysen. Die vorliegenden Aufzeichnungen bilden das Skript zur Vorlesung *B.1 Inferenzstatistik - Das Allgemeine Lineare Modell* am Institut für Psychologie der Otto-von-Guericke Universität Magdeburg im Sommersemester 2023.

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.

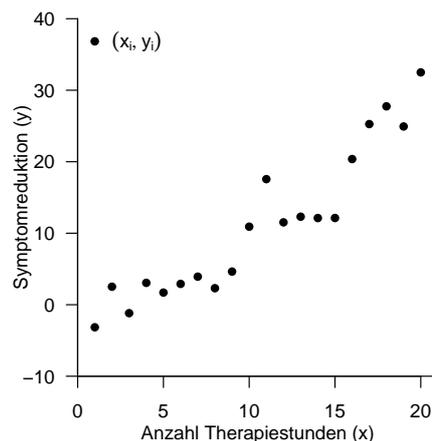


Abbildung 1.1 Beispieldatensatz

1.1 Methode der kleinsten Quadrate

Wir definieren zunächst den Begriff der *Ausgleichsgerade*.

Definition 1.1 (Ausgleichsgerade). Für $\beta := (\beta_0, \beta_1)^T \in \mathbb{R}^2$ heißt die linear-affine Funktion

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

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

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

der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_\beta(x_i)$ ihr Minimum annimmt, *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 β_0 und β_1 parameterisierte linear-affine Funktionen zusammen mit der Wertemenge des Beispieldatensatzes.

Wie bei allen linear-affinen Funktionen entspricht bei f_β der Wert von β_0 dem Wert, den f_β 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 β_0 damit dem Versatz (engl. *offset*) des Funktionsgraphen von $y = 0$ an der Stelle $x = 0$ entspricht, nennt man β_0 auch häufig den *Offsetparameter*. Analog entspricht wie bei allen linear-affinen Funktionen der Wert von β_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. β_1 enkodiert also die Stärke der Änderung der Funktionswerte pro Argumenteinheitsdifferenz und damit die Steigung (engl. *slope*) des Graphen der linear-affinen Funktion. Entsprechend wird β_1 *Steigungsparameter* oder *Slopeparameter* genannt.

Nach Definition ist die Ausgleichsgerade nun allerdings nicht eine beliebige linear-affine Funktion der Form f_β , 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 (x_i, y_i) Paaren ist der Wert dieser Summe abhängig von den Werten von β_0 und β_1 und kann deshalb durch Wahl geeigneter

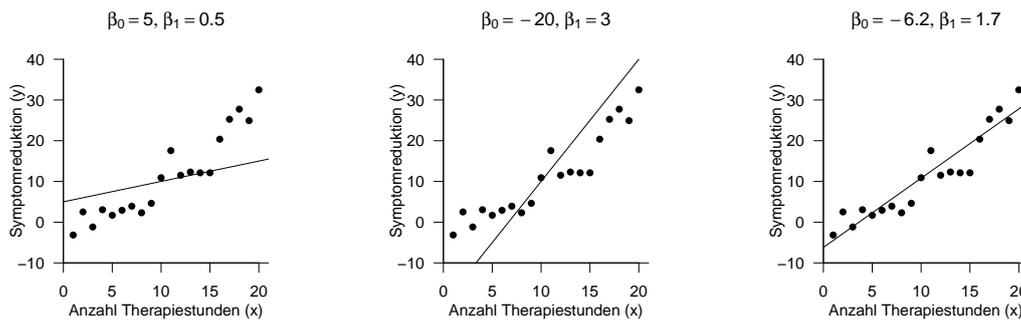


Abbildung 1.2 Linear-affine Funktionen mit unterschiedlichen Parameterwerten vor dem Hintergrund des Beispieldatensatzes

Werte von β_0 und β_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.

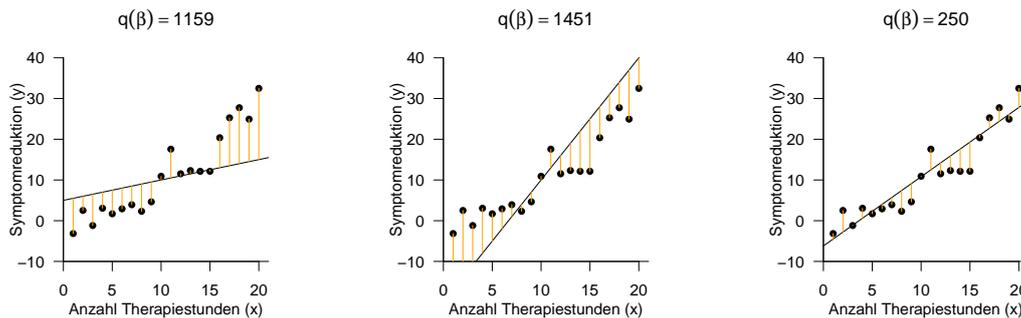


Abbildung 1.3 Vertikale Abweichungen und Quadratsummen bei unterschiedlichen Parameterwerten

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*

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

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

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

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

Minimum annimmt, die Summe der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f(x_i)$ also minimal wird. Wir betrachten 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 β_0 und β_1 und setzen diese gleich 0. Es ergibt sich zunächst

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

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 β_0 und β_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 zunächst

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

Wir halten nun zunächst fest, dass gilt

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

Weiterhin halten wir zunächst fest, dass gilt

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

In der Fortsetzung von (1.16) ergibt sich dann

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

□

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.

```

1 # Einlesen des Beispieldatensatzes
2 fname = file.path(getwd(), "_data", "401-Regression.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # Stichprobenstatistiken
6 x_bar = mean(D$x_i) # Stichprobenmittel der x_i-Werte
7 y_bar = mean(D$y_i) # Stichprobenmittel der y_i-Werte
8 s2x = var(D$x_i) # Stichprobenvarianz der x_i-Werte
9 cxy = cov(D$x_i, D$y_i) # Stichprobenkovarianz der (x_i,y_i)-Werte
10
11 # Ausgleichsgeradenparameter
12 beta_1_hat = cxy/s2x # \hat{\beta}_1, Steigungsparameter
13 beta_0_hat = y_bar - beta_1_hat*x_bar # \hat{\beta}_0, Offset Parameter
14
15 # Ausgabe
16 cat("beta_0_hat:", beta_0_hat,
17     "\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 implementiert folgender **R** Code.

```

1 # Visualisierung der Datenwerte als Punktwolke
2 plot(
3   D$x_i,
4   D$y_i,
5   pch = 16,
6   xlab = "Anzahl Therapiestunden (x)",
7   ylab = "Symptomreduktion (y)",
8   xlim = c(0, 21),
9   ylim = c(-10, 40),
10  main = TeX("\hat{\beta}_0 = -6.19, \hat{\beta}_1 = 1.66$"))
11
12 # Ausgleichsgerade
13 abline(
14   coef = c(beta_0_hat, beta_1_hat),
15   lty = 1,
16   col = "black")
17
18 # Legende
19 legend(
20   "topleft",
21   c(TeX("(x_i, y_i)$"), TeX("$f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$")),
22   lty = c(0, 1),
23   pch = c(16, NA),
24   bty = "n")

```

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

Definition 1.2 (Ausgleichspolynom). Für $\beta := (\beta_0, \dots, \beta_k)^T \in \mathbb{R}^{k+1}$ heißt die Polynomfunktion k ten Grades

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

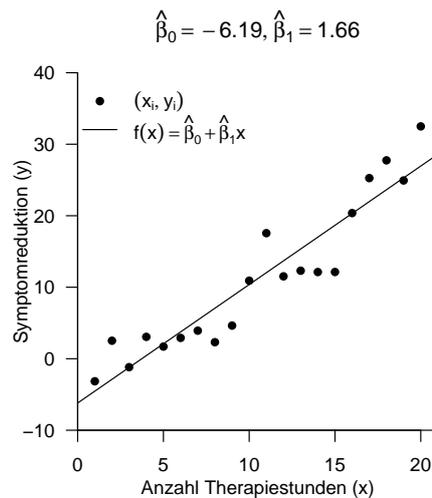


Abbildung 1.4 Ausgleichsgerade für den Beispieldatensatz

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

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

der quadrierten vertikalen Abweichungen der y_i von den Funktionswerten $f_{\beta}(x_i)$ ihr Minimum annimmt, das *Ausgleichspolynom kten 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 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.

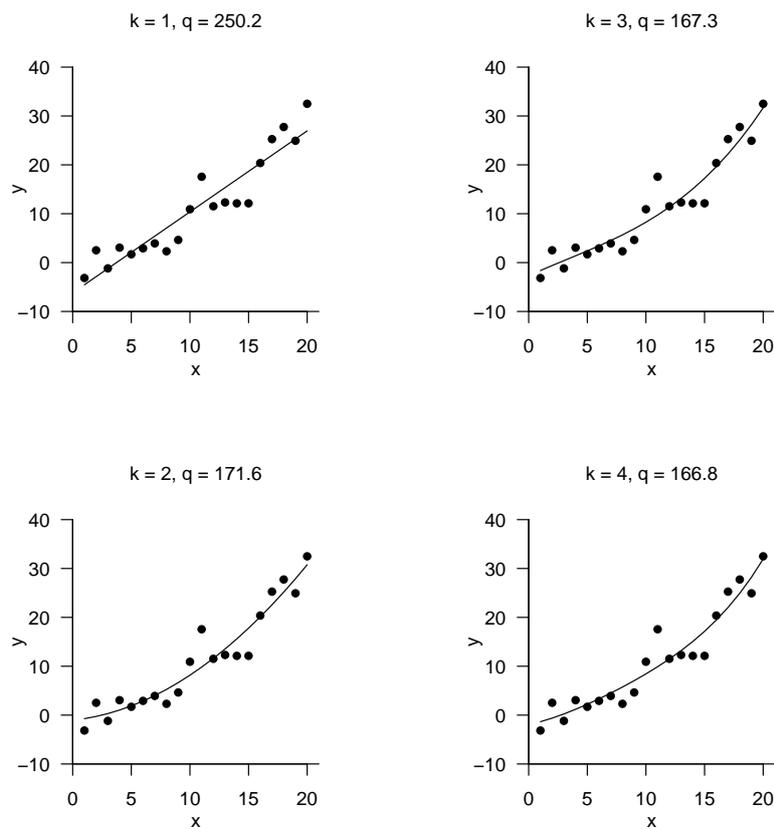


Abbildung 1.5 Ausgleichspolynome ersten bis vierten Grades für den Beispieldatensatz

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 1.3 (Modell der einfachen linearen Regression). Es sei

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

wobei

- v_i beobachtbare Zufallsvariablen sind, die Werte einer abhängigen Variable modellieren,
- $x_i \in \mathbb{R}$ fest vorgegebene *Prädiktorwerte* oder *Regressorwerte* sind, die Werte einer unabhängigen Variable modellieren,
- $\beta_0, \beta_1 \in \mathbb{R}$ wahre, aber unbekannte, Offset- und Steigungsparameterwerte sind und
- ε_i unabhängig und identisch normalverteilte nicht-beobachtbare Zufallsvariablen mit wahren, aber unbekanntem, Varianzparameter $\sigma^2 > 0$ sind, die *Fehler-* oder *Störvariablen* modellieren.

Dann heißt (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 (abhängige) Zufallsvariablen v_i anhand der Werte x_i einer unabhängigen Variable, der Parameterwerte β_0 und β_1 sowie durch Addition der normalverteilten Fehlervariablen ε_i generiert wird. Das Modell hat dabei drei Parameter, den Offsetparameter β_0 , den Steigungsparameter β_1 und den Varianzparameter σ^2 der normalverteilten Fehlervariablen. Addition der festen Werte β_0 und $\beta_1 x_i$ zu der normalverteilten Zufallsvariable ε_i impliziert dabei eine Normalverteilung von v_i . Dies ist die Aussage folgenden Theorems.

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

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

lässt sich mit $\mu_i := \beta_0 + \beta_1 x_i$ äquivalent in der Form

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

schreiben.

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

$$v_i = f(\varepsilon_i) \text{ mit } f: \mathbb{R} \rightarrow \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + (\beta_0 + \beta_1 x_i) \quad (1.25)$$

Mit dem WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

$$\begin{aligned}
 p_{v_i}(y_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{1} \right) \\
 &= N(y_i - \beta_0 - \beta_1 x_i; 0, \sigma^2) \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i - 0)^2 \right) \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (y_i - (\beta_0 + \beta_1 x_i))^2 \right) \\
 &= N(y_i; \beta_0 + \beta_1 x_i, \sigma^2).
 \end{aligned} \tag{1.26}$$

Definition von $\mu_i := \beta_0 + \beta_1 x_i$ ergibt dann die Aussage des Theorems.

□

Theorem 1.2 besagt insbesondere, dass die Datenvariablen v_i univariat normalverteilte Zufallsvariablen, 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 $\sigma^2 := 1$.

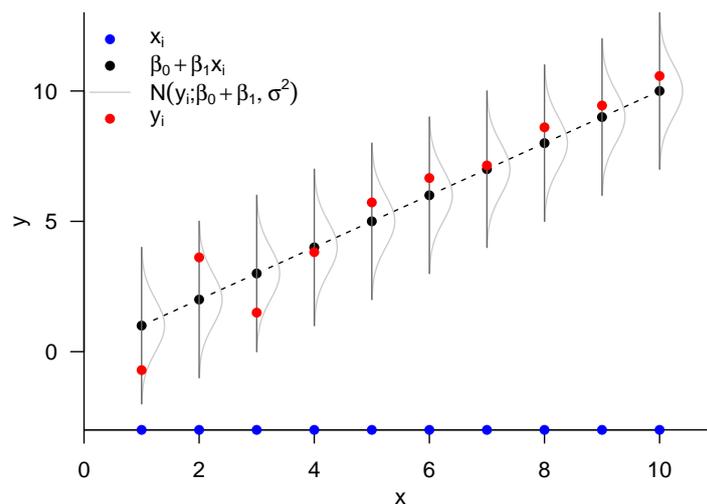


Abbildung 1.6 Modell der einfachen linearen Regression für $\beta_0 := 0, \beta_1 := 1$ und $\sigma^2 := 1$

Da es sich bei dem Modell der einfachen linearen Regression um ein parametrisches statistisches 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 Steigungsparameter mit den Werten der Ausgleichsgeradenparameter identisch sind. Dies ist eine der Aussagen folgenden Theorems. Wir verzichten hier bei den Schätzern aus Gründen der notationstechnischen Übersichtlichkeit auf ^{ML} Superskripte.

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

$$v_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \tag{1.27}$$

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

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2}, \quad \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und} \quad \hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2. \quad (1.28)$$

Beweis. Wir zeigen zunächst, dass die Ausgleichsgeradenparameter $\hat{\beta}_0$ und $\hat{\beta}_1$ den entsprechenden Maximum-Likelihood Schätzern gleichen. Dazu halten wir zunächst fest, dass aufgrund der Unabhängigkeit der v_1, \dots, v_n die Likelihood-Funktion des Modells der einfachen linearen Regression bezüglich β_0 und β_1 die Form

$$\begin{aligned} L : \mathbb{R}^2 &\rightarrow \mathbb{R}_{>0}, (\beta_0, \beta_1) \mapsto L(\beta_0, \beta_1) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2\right) \end{aligned} \quad (1.29)$$

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

$$q := \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (1.30)$$

minimal und damit $-q$ maximal wird. Im Rahmen des Beweises der Ausgleichsgeradenform haben wir aber schon gezeigt, dass der Term (1.30) für

$$\hat{\beta}_1 := \frac{c_{xy}}{s_x^2} \quad \text{und} \quad \hat{\beta}_0 := \bar{y} - \hat{\beta}_1 \bar{x} \quad (1.31)$$

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 σ^2 an der Stelle von $\hat{\beta}_0$ und $\hat{\beta}_1$. Wir erhalten die Likelihood-Funktion

$$L : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, \sigma^2 \mapsto L(\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2\right) \quad (1.32)$$

und die entsprechende Log-Likelihood-Funktion

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

In Analogie zu der Herleitung des Maximum-Likelihood Schätzers für σ^2 im Normalverteilungsmodell ergibt sich unter Beachtung von

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (1.34)$$

dann hier

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2. \quad (1.35)$$

□

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.

```
1 # Einlesen des Beispieldatensatzes
2 fname = file.path(getwd(), "_data", "401-Regression.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # Stichprobenstatistiken
6 n = length(D$y_i) # Anzahl Datenpunkte
```

```
7 x_bar = mean(D$x_i) # Stichprobenmittel der x_i-Werte
8 y_bar = mean(D$y_i) # Stichprobenmittel der y_i-Werte
9 s2x = var(D$x_i) # Stichprobenvarianz der x_i-Werte
10 cxy = cov(D$x_i, D$y_i) # Stichprobenkovarianz der (x_i,y_i)-Werte
11
12 # Steigungs- und Offsetparameterschätzung
13 beta_1_hat = cxy/s2x
14 beta_0_hat = y_bar - beta_1_hat*x_bar
15
16 # Varianzparameterschätzung
17 sigsq_hat = (1/n)*sum((D$y_i-(beta_0_hat+beta_1_hat*D$x_i))^2)
18
19
20 # Ausgabe
21 cat("beta_0_hat:" , beta_0_hat,
22     "\nbeta_1_hat:", beta_1_hat,
23     "\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 Bedeutung der Funktion der quadrierten vertikalen Abweichungen.
5. Geben Sie das Theorem zur Ausgleichsgerade wieder.
6. Skizzieren Sie den Beweis des Theorems zur Ausgleichsgeraden.
7. Definieren Sie den Begriff des Ausgleichspolynoms.
8. Erläutern Sie die Motivation des einfachen linearen Regressionsmodells.
9. Definieren Sie das Modell der einfachen linearen Regression.
10. Geben Sie das Theorem zur Datenverteilung der einfachen linearen Regression wieder.
11. Geben Sie das Theorem zur Maximum-Likelihood Schätzung der Parameter der einfachen linearen Regression an.
12. Skizzieren Sie den Beweis des Theorems zur Maximum-Likelihood Schätzung der Parameter der einfachen linearen Regression.

2 Korrelation

Die Konzepte der Regression und der 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 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 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 erste Aspekte zum Verständnis der Stichprobenkorrelation im probabilistischen Kontext. 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 ξ 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 ξ und v und $\mathbb{S}(\xi)$ und $\mathbb{S}(v)$ die Standardabweichungen von ξ und v , respektive, bezeichnen.

•

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

Liegen von zwei Zufallsvariablen Realisationen 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

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

- Die Stichprobenstandardabweichungen x_i und y_i definiert als

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

- 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 *Pearson's Stichprobenkorrelationskoeffizient* genannt.

•

Folgender R Code wertet die Stichprobenkorrelation des Beispieldatensatzes aus.

```

1 # Laden des Beispieldatensatzes
2 fname = file.path("../data/402-Korrelation.csv") # Dateipfad
3 D = read.table(fname, sep = ",", header = TRUE) # Laden als Dataframe
4 x_i = D$x_i # x_i Werte
5 y_i = D$y_i # y_i Werte
6 n = length(x_i) # n
7
8 # "Manuelle" Berechnung der Stichprobenkorrelation
9 x_bar = (1/n)*sum(x_i) # \bar{x}
10 y_bar = (1/n)*sum(y_i) # \bar{y}
11 s_x = sqrt(1/(n-1)*sum((x_i - x_bar)^2)) # s_x
12 s_y = sqrt(1/(n-1)*sum((y_i - y_bar)^2)) # s_y
13 c_xy = 1/(n-1) * sum((x_i - x_bar) * (y_i - y_bar)) # c_{xy}
14 r_xy = c_xy/(s_x * s_y) # r_{xy}
15 print(r_xy) # Ausgabe

```

```
[1] 0.9378162
```

```

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

```

```
[1] 0.9378162
```

Im Beispieldatensatz sind die Anzahl der Therapiestunden und die Symptomreduktion also mit $r_{xy} = 0.93$ 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 aber nicht zwangsläufig, dass diese Korrelation irrelevant ist (man denke an Gesundheitsrisikofaktoren), genauso wenig wie eine hohe Korrelation zweier Variablen trivial sein kann (man denke an die Korrelation von Körpergröße und Schuhgröße).

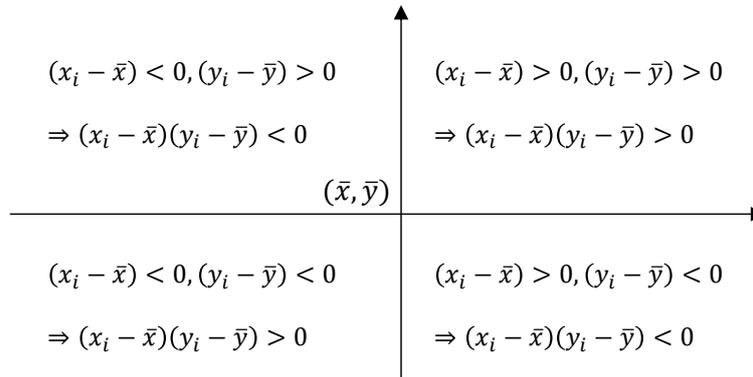


Abbildung 2.1 Mechanik der Stichprobenkorrelationssummenterme.

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.

Der Vorteil des Stichprobenkorrelationskoeffizienten gegenüber der Stichprobenkovarianz als Zusammenhangsmaß ist es, dass der absolute Wert des Stichprobenkorrelationskoeffizienten bei linear-affiner Transformation der zugrundeliegende 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.

Definition 2.3 (Stichprobenkorrelation bei linear-affiner Transformation). Für einen Datensatz $\{(x_i, y_i)\}_{i=1, \dots, n} \subset \mathbb{R}^2$ sei $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1, \dots, n} \subset \mathbb{R}^2$ der linear-affin transformierte Datensatz mit

$$(\tilde{x}_i, \tilde{y}_i) = (a_x x_i + b_x, a_y y_i + b_y), a_x, a_y \neq 0. \tag{2.6}$$

gegeben. Dann gilt

$$|r_{\tilde{x}\tilde{y}}| = |r_{xy}|. \tag{2.7}$$

•

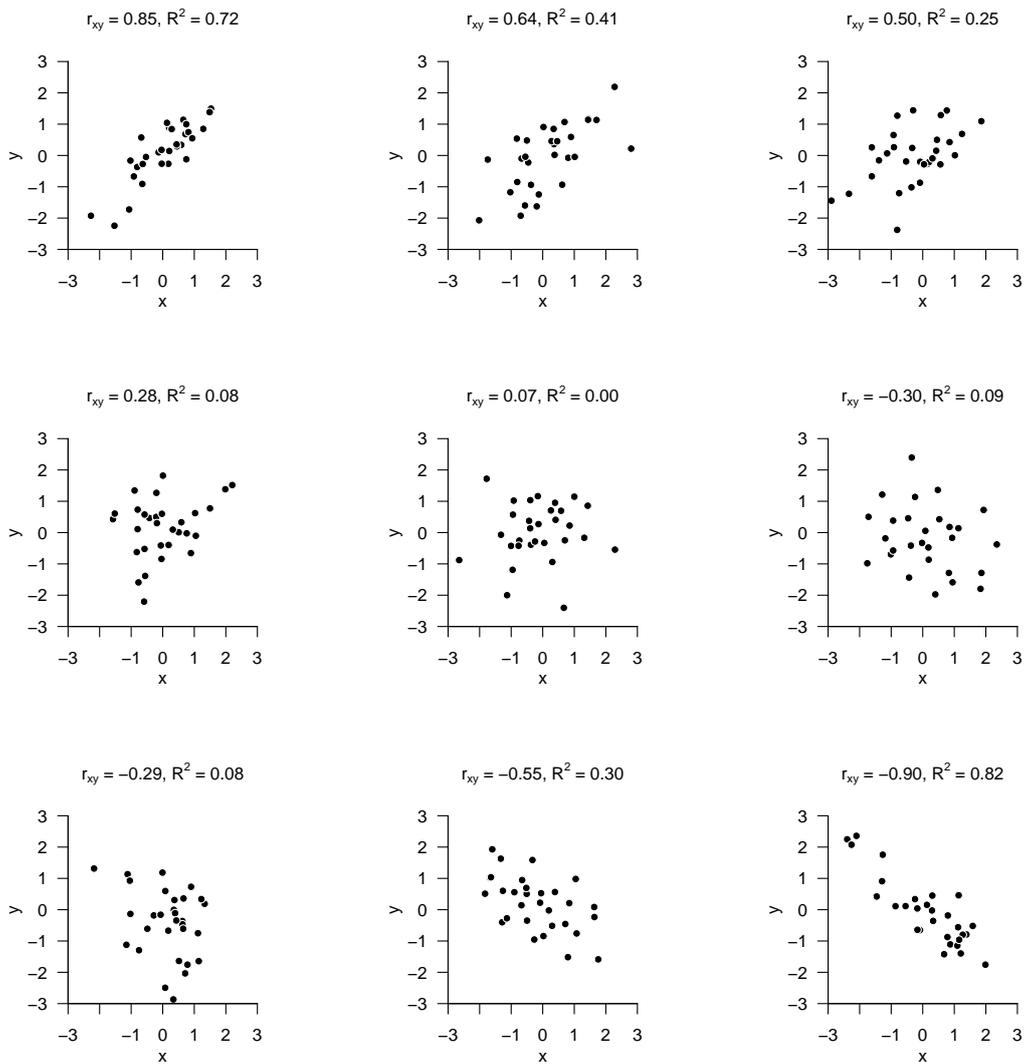


Abbildung 2.2 Beispiele bivariater Datensätze und ihrer Stichprobenkorrelationen.

Beweis.

$$\begin{aligned}
 r_{\tilde{x}\tilde{y}} &:= \frac{\frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_i - \bar{\tilde{x}})(\tilde{y}_i - \bar{\tilde{y}})}{\sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n \tilde{x}_i - \bar{\tilde{x}}\right)^2} \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n \tilde{y}_i - \bar{\tilde{y}}\right)^2}} \\
 &= \frac{\sum_{i=1}^n (a_x x_i + b_x - (a_x \bar{x} + b_x))(a_y y_i + b_y - (a_y \bar{y} + b_y))}{\sqrt{\sum_{i=1}^n (a_x x_i + b_x - (a_x \bar{x} + b_x))^2} \sqrt{\sum_{i=1}^n (a_y y_i + b_y - (a_y \bar{y} + b_y))^2}} \\
 &= \frac{a_x a_y \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{a_x^2 \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{a_y^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{2.8} \\
 &= \frac{a_x a_y}{|a_x| |a_y|} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\
 &= \frac{a_x a_y}{|a_x| |a_y|} \frac{c_{xy}}{s_x s_y} \\
 &= \frac{a_x a_y}{|a_x| |a_y|} r_{xy}.
 \end{aligned}$$

Also folgt, durch Durchspielen aller möglichen Vorzeichenfälle, dass

$$|r_{\tilde{x}\tilde{y}}| = |r_{xy}|. \tag{2.9}$$

□

2.2 Korrelation und Bestimmtheitsmaß

Das sogenannte *Bestimmtheitsmaß* R^2 ist eine beliebte, häufig reportierte Statistik zur Beschreibung der Zusammenhangsstärke der Werte einer unabhängigen und einer abhängigen Variable. Numerisch handelt es sich 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 ohne Vorteil. Wir wollen hier trotzdem etwas genauer auf R^2 eingehen, da ein tieferes Verständnis von R^2 einerseits den Einstieg in das Konzept der *Varianzzerlegungen* 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 2.4 (Erklärte Werte und Residuen einer Ausgleichsgerade). Gegeben sei ein Datensatz $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ und die zu diesem Datensatz gehörende Ausgleichsgerade

$$f_{\hat{\beta}} : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f_{\hat{\beta}}(x) := \hat{\beta}_0 + \hat{\beta}_1 x \tag{2.10}$$

Dann werden für $i = 1, \dots, n$

$$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{2.11}$$

die durch die Ausgleichsgerade *erklärten Werte* genannt und

$$\hat{\varepsilon}_i := y_i - \hat{y}_i \tag{2.12}$$

werden 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.

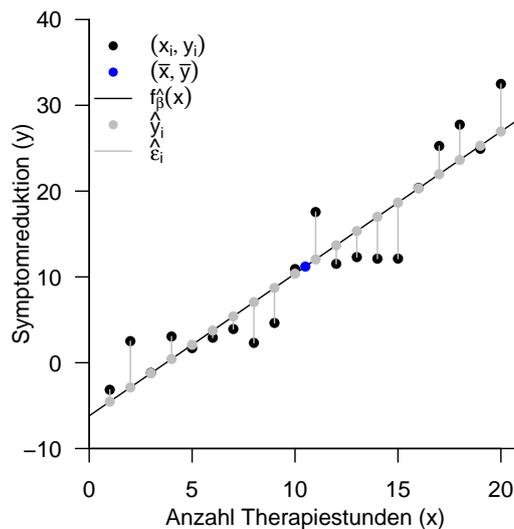


Abbildung 2.3 Ausgleichsgerade, erklärte Werte und Residuen für den Beispieldatensatz.

Mithilfe der Begriffe der erklärten Werte und Residuen lässt sich nun folgende *Quadratsummenzerlegung* beim Vorliegen einer Ausgleichsgerade eines Datensatzes 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 für*

$$\bar{y} := \frac{1}{n} \sum_{i=1}^n y_i \text{ und } \hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \text{ für } i = 1, \dots, n \tag{2.13}$$

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

- die Total Sum of Squares definiert als

$$SQT := \sum_{i=1}^n (y_i - \bar{y})^2 \tag{2.14}$$

- die Explained Sum of Squares definiert als

$$SQE := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \tag{2.15}$$

- die Residual Sum of Squares *definiert als*

$$SQR := \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.16)$$

Dann gilt

$$SQT = SQE + SQR \quad (2.17)$$

Beweis.

$$\begin{aligned} SQT &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n ((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2 \\ &= \sum_{i=1}^n ((y_i - \hat{y}_i)^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2) \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= SQE + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + SQR \\ &= SQE + SQR \end{aligned} \quad (2.18)$$

Dabei ergibt sich die letzte Gleichung mit

$$\bar{\hat{y}} := \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y} \quad (2.19)$$

und damit auch

$$\bar{\hat{y}} = \bar{y} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \Leftrightarrow \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i \Leftrightarrow \bar{y} \sum_{i=1}^n \hat{y}_i = \bar{y} \sum_{i=1}^n y_i \quad (2.20)$$

sowie

$$\bar{\hat{y}} = \bar{y} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i \Leftrightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i \Leftrightarrow \sum_{i=1}^n y_i \hat{y}_i = \sum_{i=1}^n \hat{y}_i \hat{y}_i \quad (2.21)$$

aus

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n (y_i \hat{y}_i - y_i \bar{y} - \hat{y}_i \hat{y}_i + \hat{y}_i \bar{y}) \\ &= \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n y_i \bar{y} - \sum_{i=1}^n \hat{y}_i \hat{y}_i + \sum_{i=1}^n \hat{y}_i \bar{y} \\ &= \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n \hat{y}_i \hat{y}_i + \bar{y} \sum_{i=1}^n \hat{y}_i - \bar{y} \sum_{i=1}^n y_i \\ &= 0 + 0 \\ &= 0 \end{aligned} \quad (2.22)$$

□

Die Begriffsbildungen von Theorem 2.1 erklären sich intuitiv wie folgt:

- SQT repräsentiert die Gesamtvariabilität der y_i -Werte um ihren Mittelwert \bar{y} .

- 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 - und y_i -Werte
- SQR ist die Summe der quadrierten Residuen, denn es gilt

$$\text{SQR} := \sum_{i=1}^n (y_i - \hat{y}_i)^2 := \sum_{i=1}^n \hat{\varepsilon}_i^2 \tag{2.23}$$

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 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 2.5 (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 zugehörigen Explained Sum of Squares SQE und Total Sum of Squares SQT heißt

$$R^2 := \frac{\text{SQE}}{\text{SQT}} \tag{2.24}$$

Bestimmtheitsmaß oder *Determinationskoeffizient*.

•

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

$$R^2 = r_{xy}^2. \tag{2.25}$$

Beweis. Wir halten zunächst fest, dass mit

$$\bar{\hat{y}} := \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y} \tag{2.26}$$

folgt, dass

$$\begin{aligned}
 \text{SQE} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
 &= \sum_{i=1}^n (\hat{y}_i - \tilde{y})^2 \\
 &= \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})^2 \\
 &= \sum_{i=1}^n (\hat{\beta}_1 (x_i - \bar{x}))^2 \\
 &= \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2
 \end{aligned} \tag{2.27}$$

Damit ergibt sich dann

$$\begin{aligned}
 R^2 &= \frac{\text{SQE}}{\text{SQT}} \\
 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
 &= \hat{\beta}_1^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
 &= \frac{c_{xy}^2 \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}{s_x^4 \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \\
 &= \frac{c_{xy}^2 s_x^2}{s_x^4 s_y^2} \\
 &= \frac{c_{xy}^2}{s_x^2 s_y^2} \\
 &= \left(\frac{c_{xy}}{s_x s_y} \right)^2 \\
 &= r_{xy}^2.
 \end{aligned} \tag{2.28}$$

□

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 2.5 gilt $R^2 = 0$ genau dann, wenn $\text{SQE} = 0$ 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 $\text{SQE} = \text{SQT}$ 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 die durch die Ausgleichsgerade erklärte 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, dass 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.

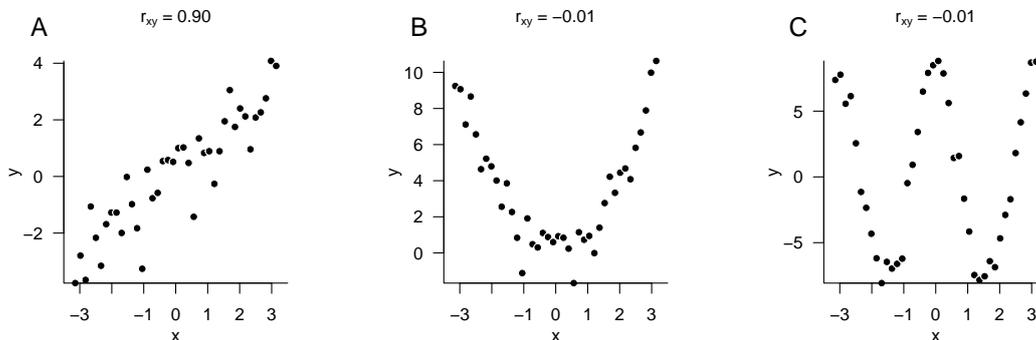


Abbildung 2.4 Korrelation und stochastische Abhängigkeit.

Abbildung 2.4 A zeigt eine Realisation des Modells

$$v_i = x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, 1) \text{ für } i = 1, \dots, n. \tag{2.29}$$

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 v_i Realisation y_i vom Wert x_i : je höher der Wert von x_i , desto höher der Erwartungswert für den Wert von v_i . Abbildung 2.4 B zeigt eine Realisation des Modells

$$v_i = x_i^2 + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, 1) \text{ für } i = 1, \dots, n. \tag{2.30}$$

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 v_i 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 v_i , es besteht ein quadratischer Zusammenhang. Ein ähnliches Bild ergibt sich beim Betrachten von Abbildung 2.4 C einer Realisation des Modells

$$v_i = 8 \cos(2x_i) + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, 1) \text{ für } i = 1, \dots, n. \tag{2.31}$$

Auch hier ergibt sich das Bild einer Abhängigkeit des Wertes der v_i 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). *ξ und v seien zwei Zufallsvariablen mit positiver Varianz. Dann besteht genau dann eine linear-affine Abhängigkeit der Form*

$$v = \beta_0 + \beta_1 \xi \text{ mit } \beta_0, \beta_1 \in \mathbb{R} \tag{2.32}$$

zwischen ξ und v , wenn

$$\rho(\xi, v) = 1 \text{ oder } \rho(\xi, v) = -1 \quad (2.33)$$

gilt.

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 den Eigenschaften von Erwartungswert und Varianz gilt, dass

$$\mathbb{E}(v) = \beta_0 + \beta_1 \mathbb{E}(\xi) \text{ und } \mathbb{V}(v) = \beta_1^2 \mathbb{V}(\xi). \quad (2.34)$$

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

$$\beta_1 > 0 \Rightarrow \mathbb{S}(v) = \beta_1 \mathbb{S}(\xi) > 0 \text{ und } \beta_1 < 0 \Rightarrow \mathbb{S}(v) = -\beta_1 \mathbb{S}(\xi) > 0. \quad (2.35)$$

Weiterhin gilt

$$\begin{aligned} v - \mathbb{E}(v) &= \beta_0 + \beta_1 \xi - \mathbb{E}(v) \\ &= \beta_0 + \beta_1 \xi - \beta_0 - \beta_1 \mathbb{E}(\xi) \\ &= \beta_1 \xi - \beta_1 \mathbb{E}(\xi) \\ &= \beta_1 (\xi - \mathbb{E}(\xi)). \end{aligned} \quad (2.36)$$

Für die Kovarianz von ξ und v ergibt sich also

$$\begin{aligned} \mathbb{C}(\xi, v) &= \mathbb{E}((v - \mathbb{E}(v))(\xi - \mathbb{E}(\xi))) \\ &= \mathbb{E}(\beta_1 (\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))) \\ &= \beta_1 \mathbb{E}((\xi - \mathbb{E}(\xi))^2) \\ &= \beta_1 \mathbb{V}(\xi). \end{aligned} \quad (2.37)$$

Damit ergibt für die Korrelation von ξ und v

$$\rho(\xi, v) = \frac{\mathbb{C}(\xi, v)}{\mathbb{S}(\xi)\mathbb{S}(v)} = \pm \frac{\beta_1 \mathbb{V}(\xi)}{\mathbb{S}(\xi)\beta_1 \mathbb{S}(\xi)} = \pm \frac{\beta_1 \mathbb{V}(\xi)}{\beta_1 \mathbb{V}(\xi)} = \pm 1. \quad (2.38)$$

□

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 von v von ξ auch immer die linear-affine Abhängigkeit von ξ von v , denn

$$v = \beta_0 + \beta_1 \xi \Leftrightarrow -\beta_0 + v = \beta_1 \xi \Leftrightarrow \xi = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} v \Leftrightarrow \xi = \tilde{\beta}_0 + \tilde{\beta}_1 v \quad (2.39)$$

mit

$$\tilde{\beta}_0 = -\frac{\beta_0}{\beta_1} \text{ und } \tilde{\beta}_1 = \frac{1}{\beta_1}. \quad (2.40)$$

2.4 Literaturhinweise

Der Begriff der Korrelation erscheint, allerdings basierend auf früheren Arbeiten z.B. von Bravais (1844), zunächst bei Galton (1890) (vgl. Stigler (1986)) und wird u.a. 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, Stichprobenstandardabweichung, Stichprobenkovarianz und Stichprobenkorrelation 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 das Theorem zur Stichprobenkorrelation bei linear-affinen Transformationen wieder.
5. Erläutern Sie das Theorem zur Stichprobenkorrelation bei linear-affinen Transformationen.
6. Geben Sie die Definitionen von erklärten Werten und Residuen einer Ausgleichsgerade wieder.
7. Geben Sie das Theorem zur Quadratsummenzerlegung bei einer Ausgleichsgerade wieder.
8. Erläutern Sie die intuitiven Bedeutungen von SQT, SQE und SQR.
9. Geben Sie die Definition des Bestimmtheitsmaßes R^2 wieder.
10. Geben Sie das Theorem zum Zusammenhang von Stichprobenkorrelation und Bestimmtheitsmaß wieder.
11. Erläutern Sie die Bedeutung von hohen und niedrigen R^2 Werten im Lichte der Ausgleichsgerade.
12. Geben Sie das Theorem zum Zusammenhang von Korrelation und linear-affiner Abhängigkeit wieder.

3 Matrizen

Matrizen sind die Worte 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 weiterhin zur Repräsentation linearer Abbildungen und von Vektorräumen, hier werden Vektoren dann als spezielle Matrizen aufgefasst. In diesem Abschnitt geben wir eine erste Einführung zum Umgang mit Matrizen, wobei wir auf abstrakte Konzepte der linearen Algebra im Wesentlichen verzichten. Wir führen zunächst den Matrixbegriff ein und diskutieren dann mit der Matrixaddition, Matrixsubtraktion, Skalarmultiplikation und der Matrixtransposition erste grundlegende Operationen. Wir führen dann die zentralen Begriffe der Matrixmultiplikation und der Matrixinversion ein. Mit der Matrixdeterminante diskutieren wir dann eine erste Maßzahl zur Beschreibung von Matrizen und schließen 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 indiziert. Zum Beispiel gilt für

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

dass $a_{32} = 4$. 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 die 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} , 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 **R** Vektoren mithilfe der `matrix()` Funktion in die Repräsentation einer mathematischen Matrix transformiert werden. Die Einträge eines **R** Vektors werden dabei anhand der spezifizierten Zeilenanzahl `nrow` anhand 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

```
1 # Spaltenweise Definition von A (R default)
2 A = matrix(c(2,1,3,6,0,5), nrow = 2)
3 print(A)
```

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

R folgt hier per default einer sogenannten *column-major-order*, d.h. die Elemente des **R** 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 **R** 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 **R** Vektors gefüllt wird und dann die zweite Zeile usw. bis alle Elemente des Vektors auf die Matrix verteilt sind.

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

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

```
1 # Zeilenweise Definition von B
2 B = matrix(c(4,1,0,
```

```

3         -4,2,0),
4         nrow = 2,
5         byrow = TRUE)
6 print(B)

```

```

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

```

3.2 Grundlegende Matrixoperationen

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 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 Additionsoperation elementweise definiert ist.

Beispiel

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

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

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

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

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

```

1 # Definition
2 A = matrix(c(2, -3, 0,
3             1, 6, 5),
4             nrow = 2,
5             byrow = TRUE)
6 B = matrix(c(4, 1, 0,
7             -4, 2, 0),
8             nrow = 2,
9             byrow = TRUE)
10
11 # Addition
12 C = A + B
13 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. Es seien $A, B \in \mathbb{R}^{n \times m}$. Dann ist die *Subtraktion* von A und B definiert als die Abbildung

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

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 **R** führt man diese Rechnung wie folgt aus.

```

1 # Subtraktion
2 D = A - B
3 print(D)

```

```

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

```

3.2.3 Skalarmultiplikation

Die *Skalarmultiplikation* einer Matrix bezeichnet die Multiplikation eines Skalars mit einer Matrix.

Definition 3.4. 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

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



Die Skalarmultiplikation ist mit dieser Definition also elementweise definiert.

Beispiel

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

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

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 \mathbf{R} führt man diese Skalarmultiplikation aus wie folgt.

```

1 # Definitionen
2 A = matrix(c(3,1,1,
3             5,2,5,
4             2,7,1,
5             3,4,2),
6           nrow = 4,
7           byrow = TRUE)
8 c = -3
9
10 # Skalarmultiplikation
11 B = c*A
12 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. *Das Tripel $(\mathbb{R}^{n \times m}, +, \cdot)$ mit der oben definierten Matrixaddition und Skalarmultiplikation ist ein Vektorraum. Insbesondere gelten damit 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, welcher sich mit einigem Notationsaufwand direkt aus dem elementweisen Charakter von Matrixaddition und Skalarmultiplikation sowie aus dem Umgang mit den reellen Zahlen bekannten Rechenregeln ergibt. Das im Theorem erwähnte neutrale Element der Addition wird *Nullmatrix* genannt, wir werden dazu später eine allgemeine Notation einführen. Die inversen Elemente der Addition sind durch

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

gegeben und erlauben, 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:

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

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

mit

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

•

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

$$A^T = A. \tag{3.19}$$

(2) Es gilt

$$(A^T)^T = A. \tag{3.20}$$

(3) Es gilt

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

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

$$A := \begin{pmatrix} 2 & 3 & 0 \\ 1 & 6 & 5 \end{pmatrix}, \tag{3.22}$$

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

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

In **R** führt man die Transposition einer Matrix wie folgt durch.

```

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

```

```

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

```

```

1 # Transposition
2 AT = t(A)
3 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 klarmacht.

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

$$(A + B)^T = A^T + B^T. \quad (3.25)$$

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

$$(A - B)^T = A^T - B^T. \quad (3.26)$$

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

$$(cA)^T = cA^T. \quad (3.27)$$

3.3 Matrixmultiplikation

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

Definition 3.6. 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

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

•

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 der Zeile aus A ein korrespondierendes Element in der Spalte von B .
- (3) Man multipliziert die korrespondierenden Elemente miteinander.
- (4) Die Summe dieser Produkte ist dann der Eintrag mit Index ij in AB .

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 **R** nutzt man für die Matrixmultiplikation den `%%` Operator.

```

1 # Definitionen
2 A = matrix(c(2,-3,0,
3             1, 6,5),
4             nrow = 2,
5             byrow = TRUE)
6 B = matrix(c( 4,2,
7             -1,0,
8             1,3),
9             nrow = 3,
10            byrow = TRUE)
11
12 # Matrixmultiplikation
13 C = A %% B
14 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 **R** überprüft man diese Rechnung wie folgt.

```

1 # Definitionen
2 A = matrix(c(2,-3,0,
3             1, 6,5),
4             nrow = 2,
5             byrow = TRUE)
6 B = matrix(c( 4,2,
7             -1,0,
8             1,3),
9             nrow = 3,
10            byrow = TRUE)
11
12 # Matrixmultiplikation
13 D = B %*% A
14 print(D)

```

```

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

```

Ist allerdings eine Matrixmultiplikation aufgrund nicht-adäquater Matrizengrößen nicht definiert, so lässt sich diese auch nicht numerisch auswerten.

```

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

```

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

Folgendes Theorem, das wir nicht beweisen wollen, stellt den Bezug zwischen dem Skalarprodukt zweier Vektoren und der Multiplikation zweier Matrizen her. Dieser ergibt sich im Wesentlichen durch die Identifikation von \mathbb{R}^n und $\mathbb{R}^{n \times 1}$ und 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.

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

$$\langle x, y \rangle = x^T y. \quad (3.36)$$

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

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

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

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

die Spalten von B , also

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

Dann gilt

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

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 einschlägige 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*

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

(2) *Die Kombination von Matrizenmultiplikation 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 $A(cB)$ anhand von

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

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

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

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

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

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

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

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

Dann gilt

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = BA \quad (3.48)$$

□

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

$$(AB)^T = B^T A^T. \quad (3.49)$$

Beweis. Ein Beweis ergibt sich wie folgt

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

□

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

$$Ax = b. \quad (3.51)$$

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

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

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

$$Ax = b \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \Leftrightarrow \begin{cases} 1x_1 + 2x_2 = 5 \\ 3x_1 + 4x_2 = 11 \end{cases} \quad (3.53)$$

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 1×1 Matrix, also ein Skalar, sei und ebenso x und b , dass wir also für $a, x, b \in \mathbb{R}$ die Gleichung

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

haben. Um diese Gleichung nach x aufzulösen würde man natürlich beide Seiten der Gleichung mit dem *multiplikativem 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 gelten

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

Ganz konkret etwa

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

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

$$A^{-1}A = "1". \tag{3.58}$$

resultiert. Dann hätte man nämlich

$$Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b \tag{3.59}$$

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

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

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, d.h. es gilt für $A \in \mathbb{R}^{n \times m}$, dass

$$I_n A = A \text{ und } A I_m = A. \tag{3.61}$$

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 $1 \leq j \leq n$

$$d_{ij} = 0 \cdot a_{1j} + 0 \cdot a_{2j} + \dots + 0 \cdot a_{i-1,j} + 1 \cdot a_{ij} + \dots + 0 \cdot a_{i+1,j} + 0 \cdot a_{nj} = a_{ij} \tag{3.62}$$

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, so dass

$$A^{-1}A = AA^{-1} = I_n \tag{3.63}$$

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 also Lösungen haben, müssen es aber nicht). 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

$$A := \begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix} \quad (3.64)$$

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

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

mit

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

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

Zum Berechnen inverser Matrizen

2×2 bis etwa 5×5 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 **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 **R** Code.

```
1 # Definition
2 A = matrix(c(2,1,
```

```

3           3,4),
4           nrow = 2,
5           byrow = TRUE)
6
7 # Berechnen von A^{-1}
8 print(solve(A))

```

```

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

```

```

1 # Überprüfen der Eigenschaften einer inversen Matrix
2 print(solve(A) %*% A)

```

```

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

```

```

1 # Bei der umgekehrten Berechnung ergeben sich kleine Rundungsfehler
2 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.

```

1 B = matrix(c(1,0,
2             0,0),
3           nrow = 2,
4           byrow = 2)
5 solve(B)

```

```
Error in solve.default(B): Lapackroutine dgesv: System ist genau 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

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

d.h. 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

$$|A| := a_{11} \quad \text{für } n = 1 \quad (3.71)$$

$$|A| := \sum_{j=1}^n a_{1j}(-1)^{1+j} \det(A_{1j}) \quad \text{für } n > 1 \quad (3.72)$$

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 1×1 Matrix zurück, die durch ihr einziges Element gegeben ist. Für

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \quad (3.73)$$

ergeben sich dabei zum Beispiel folgende Matrizen der Form $A_{ij} \in \mathbb{R}^{(3-1) \times (3-1)}$:

$$A_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, A_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, A_{21} = \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}, A_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}. \quad (3.74)$$

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

$$|A| = a_{11}a_{22} - a_{12}a_{21}. \quad (3.75)$$

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

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \quad (3.76)$$

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

$$\begin{aligned} |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\ &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| \\ &= a_{11}|(a_{22})| - a_{12}|(a_{21})| \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned} \quad (3.77)$$

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

$$\begin{aligned}
 |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\
 &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| + a_{13}(-1)^{1+3}|A_{13}| \\
 &= a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| \\
 &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
 &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\
 &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\
 &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}.
 \end{aligned} \tag{3.78}$$

□

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

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

Dabei bezieht sich die Merkregeln bei 3×3 Matrizen auf das Schema

$$\left(\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \right). \tag{3.79}$$

Beispiele für Determinanten von 2×2 und 3×3 Matrizen

Es seien

$$A := \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ und } C := \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \tag{3.80}$$

Dann ergeben sich

$$|A| = 2 \cdot 4 - 1 \cdot 3 = 8 - 3 = 5 \tag{3.81}$$

und

$$|B| = 1 \cdot 0 - 0 \cdot 0 = 0 - 0 = 0 \tag{3.82}$$

und

$$|C| = 2 \cdot 1 \cdot 3 + 0 \cdot 0 \cdot 0 + 0 \cdot 0 \cdot 0 - 0 \cdot 0 \cdot 3 - 0 \cdot 0 \cdot 0 - 0 \cdot 1 \cdot 0 = 2 \cdot 1 \cdot 3 = 6. \tag{3.83}$$

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

```

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

```

[1] 5

```

1 # Matrixdefinition und Determinantenberechnung
2 B = matrix(c(1,0,
3             0,0),
4             nrow = 2,
5             byrow = TRUE)
6 det(B)

```

[1] 0

```

1 # Matrixdefinition und Determinantenberechnung
2 C = matrix(c(2,0,0,
3             0,1,0,
4             0,0,3),
5             nrow = 3,
6             byrow = TRUE)
7 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

$$|AB| = |A||B|. \quad (3.84)$$

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

$$|A| = |A^T|. \quad (3.85)$$

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

$$|A^{-1}| = \frac{1}{|A|}. \quad (3.86)$$

(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

$$|A| = \prod_{i=1}^n a_{ii}. \quad (3.87)$$

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

$$A \text{ ist invertierbar} \Leftrightarrow |A| \neq 0 \text{ und } A \text{ ist nicht invertierbar} \Leftrightarrow |A| = 0. \quad (3.88)$$

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

$$|AB| = |A||B| = |I_n| = 1. \tag{3.89}$$

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

□

Visuelle Intuition

Der abstrakte Begriff der Determinante einer quadratischen Matrix kann mithilfe des Vektorraumbegriff 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 gilt (wie wir nicht beweisen wollen), dass $|A|$ dem signierten Volumen des von $a_1, \dots, a_n \in \mathbb{R}^n$ aufgespannten Parallelotops entspricht. Um dies visuell zu veranschaulichen betrachten wir die Matrizen

$$A_1 = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, A_3 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \tag{3.90}$$

mit den jeweiligen Determinanten

$$|A_1| = 3 \cdot 2 - 1 \cdot 1 = 5, \quad |A_2| = 2 \cdot 2 - 0 \cdot 0 = 4, \quad |A_3| = 2 \cdot 2 - 2 \cdot 2 = 0. \tag{3.91}$$

Abbildung 3.1 visualisiert die entsprechende Intuition.

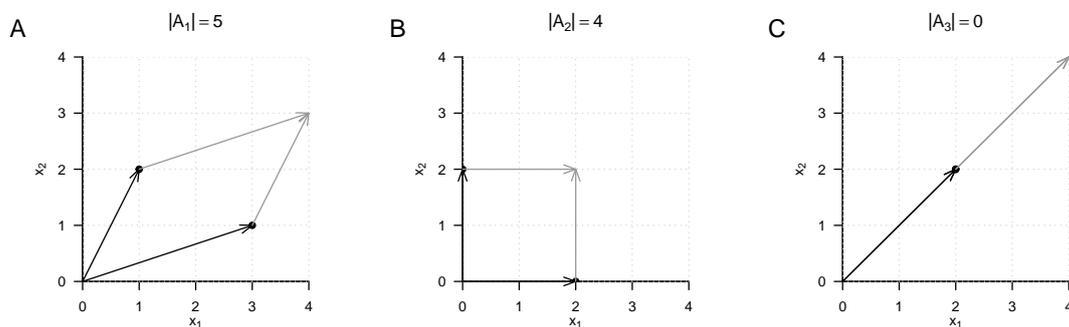


Abbildung 3.1 Determinanten als Parallelotopvolumina.

3.6 Spezielle Matrizen

In dieser Sektion stellen wir einige häufig auftretende Typen von Matrizen und ihre Eigenschaften zusammen. Zum Beweis der allermeisten Eigenschaften verweisen wir dabei auf die einschlägige Literatur.

3.6.1 Einheitsmatrizen

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

Definition 3.10 (Einheitsmatrix und Einheitsvektoren). Wir bezeichnen die *Einheitsmatrix* mit

$$I_n := (i_{jk})_{1 \leq j \leq n, 1 \leq k \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ f\u00fcr } j = k \text{ und } i_{jk} = 0 \text{ f\u00fcr } j \neq k. \quad (3.92)$$

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

$$e_i := (e_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{ij} = 1 \text{ f\u00fcr } i = j \text{ und } e_{ij} = 0 \text{ f\u00fcr } i \neq j. \quad (3.93)$$

•

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

$$I_n = (e_1 \ \dots \ e_n) \in \mathbb{R}^{n \times n} \quad (3.94)$$

F\u00fcr $n = 3$ gilt also zum Beispiel

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ und } e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.95)$$

Weiterhin gelten f\u00fcr die Einheitsvektoren bekanntlich f\u00fcr $1 \leq i, j \leq n$

$$e_i^T e_j = 0 \text{ f\u00fcr } i \neq j, e_i^T e_i = 1 \text{ und } e_i^T v = v^T e_i = v_i \text{ f\u00fcr } v \in \mathbb{R}^n. \quad (3.96)$$

3.6.2 Einsmatrizen und Nullmatrizen

Definition 3.11 (Nullmatrizen, Nullvektoren, Einsmatrizen, Einsvektoren). Wir bezeichnen *Nullmatrizen* und *Nullvektoren* mit

$$0_{nm} := (0)_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n \quad (3.97)$$

Wir bezeichnen *Einsmatrizen* und *Einsvektoren* mit

$$1_{nm} := (1)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 1_n := (1)_{1 \leq i \leq n} \in \mathbb{R}^n \quad (3.98)$$

•

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

$$0_{32} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, 0_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, 1_{32} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \text{ und } 1_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad (3.99)$$

Weiterhin gelten zum Beispiel

$$0_n 0_n^T = 0_{nn} \text{ und } 1_n 1_n^T = 1_{nn}. \quad (3.100)$$

wovon man sich durch Nachrechnen \u00fcberzeugt.

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 quadratische Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ mit den Diagonalelementen $d_1, \dots, d_n \in \mathbb{R}$ schreibt man auch als

$$D = \text{diag}(d_1, \dots, d_n). \quad (3.101)$$

Zum Beispiel gelten

$$D := \text{diag}(1, 2, 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad (3.102)$$

und für $\sigma^2 \in \mathbb{R}$

$$\Sigma = \text{diag}(\sigma^2, \sigma^2, \sigma^2) = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_3. \quad (3.103)$$

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

$$|D| = \prod_{i=1}^n d_i. \quad (3.104)$$

3.6.4 Symmetrische Matrizen

Symmetrische Matrizen sind quadratische Matrizen, die bei Transposition unverändert bleiben:

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

•

Ein Beispiel für eine symmetrische Matrix ist

$$S := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}. \quad (3.105)$$

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

Theorem 3.12 (Eigenschaften symmetrischer Matrizen).

(Summation.) $S_1 \in \mathbb{R}^{n \times n}$ und $S_2 \in \mathbb{R}^{n \times n}$ seien symmetrische Matrizen. Dann gilt

$$S_1 + S_2 = (S_1 + S_2)^T. \quad (3.106)$$

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

$$(S^{-1})^T = S^{-1}. \quad (3.107)$$

3.6.5 Positiv-definite Matrizen

Positiv-definite Matrizen sind für die probabilistische Modellierung mit multivariaten Normalverteilungen zentral.

Definition 3.14. 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.

•

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

Theorem 3.13 (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

$$A := \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, B := \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix}, \text{ und } c := 2 \quad (3.108)$$

Berechnen Sie

$$D := c(A - B^T) \text{ und } E := (cA)^T + B. \quad (3.109)$$

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:

$$ABC, ABC^T, A^T C B^T, BAC. \quad (3.110)$$

9. Es seien

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 2 & 0 \end{pmatrix} B := \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 0 & 0 \end{pmatrix} \text{ und } C := \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}. \quad (3.111)$$

Berechnen Sie die Matrixprodukte

$$AB, B^T A^T, (B^T A^T)^T, AC. \quad (3.112)$$

10. Definieren Sie die Begriff der inversen Matrix und der Invertierbarkeit einer Matrix.

11. Geben Sie die Formel für die Determinante von $A := (A_{ij})_{1 \leq i, j \leq 2} \in \mathbb{R}^2$ wieder.

12. Geben Sie die Formel für die Determinante von $A := (A_{ij})_{1 \leq i, j \leq 3} \in \mathbb{R}^3$ wieder.

13. Berechnen Sie die Determinanten von

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

14. Geben Sie die Definitionen von Einheitsmatrix und Einheitsvektoren wieder.

15. Geben Sie die Definitionen von Nullmatrizen und Einsmatrizen wieder.

16. Geben Sie die Definition einer symmetrischen Matrix wieder.

17. Geben Sie die Definition einer Diagonalmatrix wieder.

18. 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 liegt bekanntermaßen im Zentralen Grenzwertsatz: In probabilistischen Modellen repräsentieren probabilistische Terme die Summation sehr vieler Prozesse, die durch die deterministischen Bestandteile des jeweiligen Modells, also eine formalisierte wissenschaftliche Theorie, nicht erklärt werden. Nach dem Zentralen Grenzwertsatz ist die Summe dieser nicht erklärten Prozesse 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, den Generalisierungen des Allgemeinen Modells zu Hierarchischen Linearen Modellen oder Multivariaten Allgemeinen Linearen Modellen, der Bayesianischen Inferenz bei Normalverteilungsannahmen und nicht zuletzt der Theorie probabilistischer Filter, wie zum Beispiel dem Kalman-Bucy Filter.

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

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

•

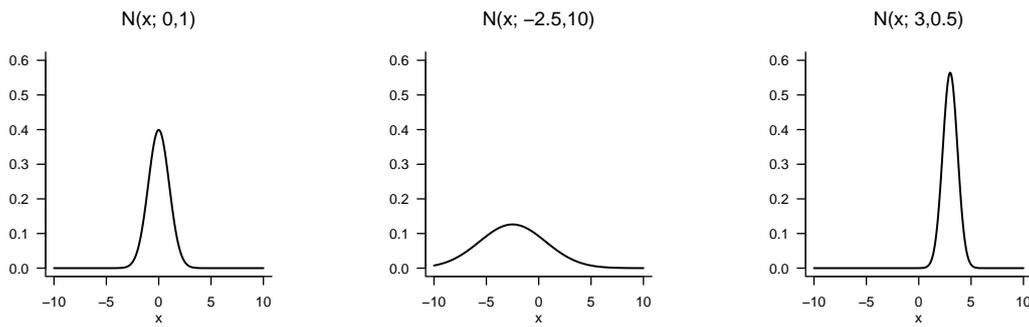


Abbildung 4.1 Wahrscheinlichkeitsdichtefunktionen univariater Normalverteilungen.

Visuell entspricht der Parameter μ einer normalverteilten Zufallsvariable dem Wert höchster Wahrscheinlichkeitsdichte und der Parameter σ^2 spezifiziert die Breite der WDF (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 ξ_1 und ξ_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 $\mu \in \mathbb{R}^2$ und $\Sigma \in \mathbb{R}^{2 \times 2}$ 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

Folgender **R** Code zeichnet das obige Theorem anhand konkreter Beispielwerte für $\mu_1, \mu_2, \sigma_1, \sigma_2$ und ρ nach und gibt die Parameter μ und Σ der resultierenden bivariaten WDF aus.

```

1 # Parameterdefinitionen
2 mu_1 = 5.0 # \mu_1
3 mu_2 = 4.0 # \mu_2
4 sig_1 = 1.5 # \sigma_1
5 sig_2 = 1.0 # \sigma_2
6 rho = 0.9 # \rho
7
8 # Realisierungen der standardnormalverteilten ZVen
9 n = 100 # Anzahl Realisierungen
10 zeta_1 = rnorm(n) # \zeta_1 \sim N(0,1)
11 zeta_2 = rnorm(n) # \zeta_2 \sim N(0,1)
12
13 # Evaluation von Realisierungen von \xi_1 und \xi_2
14 xi_1 = sig_1*zeta_1 + mu_1 # Realisierungen von zeta_1
15 xi_2 = sig_2*(rho*zeta_1 + sqrt(1-rho^2)*zeta_2) + mu_2 # Realisierungen von zeta_2
16
17 # Parameter der gemeinsamen Verteilung von \xi_1 und \xi_2
18 mu = matrix(c(mu_1, # \mu \in \mathbb{R}^2
19              mu_2),
20             nrow = 2, byrow = TRUE)
21 Sigma = matrix(c(sig_1^2, rho*sig_1*sig_2, # \Sigma \in \mathbb{R}^{2 \times 2}
22                rho*sig_1*sig_2, sig_2^2),
23               nrow = 2, byrow = TRUE)
24 print(mu)

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

1 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 postulierten WDF sind in Abbildung 4.2 dargestellt.

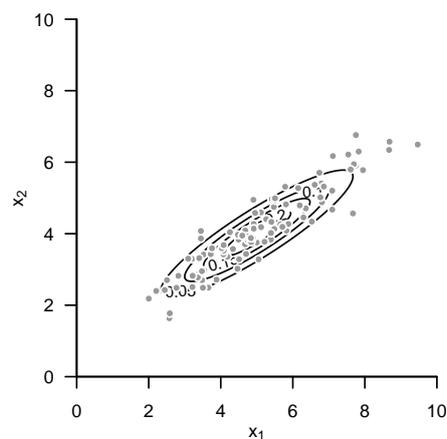


Abbildung 4.2 Konstruktion bivariater Normalverteilungen.

4.2 Definition

In diesem Abschnitt wollen wir die multivariate Normalverteilung nun formal einführen und erste Eigenschaften angeben. Wir nutzen folgende Definition.

Definition 4.2. ξ 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 sagt man, dass ξ 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 nennt ξ 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 A,B,C zeigen die Isokonturen der WDFen bivariat normalverteilter Zufallsvektoren für $\mu = (1, 1)^T$ und

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

respektive. Abbildung 4.4 A-C zeigen jeweils 200 Realisierungen der entsprechenden Zufallsvektoren.

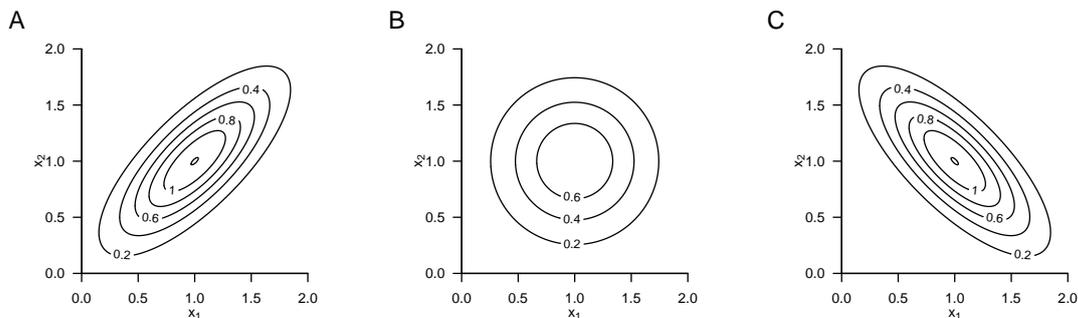


Abbildung 4.3 Wahrscheinlichkeitsdichtefunktionen bivariater Normalverteilungen.

Ohne Beweis halten wir fest, dass wie im Fall einer univariat normalverteilten Zufallsvariable, der Erwartungswert und die Kovarianzmatrix eines 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

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

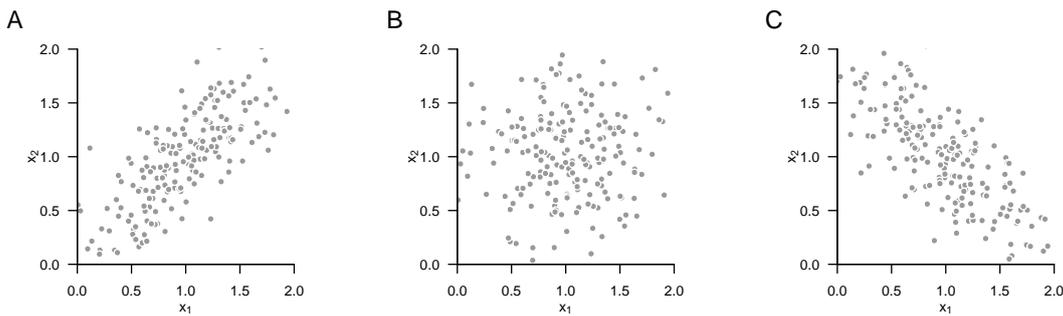


Abbildung 4.4 Realisierungen bivariat normalverteilter Zufallsvektoren.

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 ξ_1, \dots, ξ_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 ξ_i und ξ_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 (Invertierbare lineare Transformation eines normalverteilten Zufallsvektors). $\xi \sim N(\mu_\xi, \Sigma_\xi)$ sei ein normalverteilter n -dimensionaler Zufallsvektor und es sei $\zeta := A\xi$ mit einer invertierbaren Matrix $A \in \mathbb{R}^{n \times n}$. Dann gilt

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu_\xi \text{ und } \Sigma_\zeta = A\Sigma_\xi A^T. \quad (4.11)$$

Nach Theorem 4.3 ergibt die invertierbare lineare Transformation eines multivariat normalverteilten Zufallsvektors also wiederum einen multivariat normalverteilten Zufallsvektor und die Parameter der Verteilung dieses normalverteilten Zufallsvektors ergeben sich aus den Parametern der Verteilung des ursprünglichen Zufallsvektors und der Transformationsmatrix.

Beispiel

Als Beispiel betrachten wir die invertierbare lineare Transformation eines bivariaten normalverteilten Zufallsvektors ξ . Es seien

$$\mu_\xi := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ und } \Sigma_\xi := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix} \quad (4.12)$$

der Erwartungswert- und Kovarianzmatrixparameter von ξ , respektive, und es sei

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

die Transformationsmatrix. Da $|A| = -3 \neq 0$ ist A invertierbar und es gilt nach Theorem 4.3, dass

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu_\xi = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ und } A\Sigma_\xi A^T = \begin{pmatrix} 0.40 & 0.05 \\ 0.05 & 0.40 \end{pmatrix}. \quad (4.14)$$

Abbildung 4.5 A zeigt Isokonturen der WDF von ξ und Realisierungen $x^{(i)} \in \mathbb{R}^2$ von ξ für $i = 1, \dots, 50$. Abbildung 4.5 B zeigt die transformierten Realisierungen $z^{(i)} = Ax^{(i)} \in \mathbb{R}^2$ von ζ sowie die Isokonturen der WDF von ζ nach Theorem 4.3.

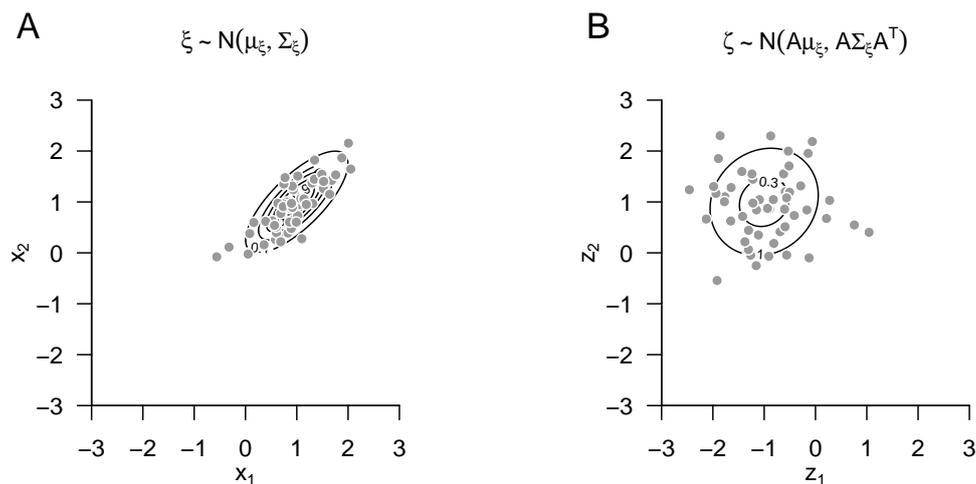


Abbildung 4.5 invertierbare lineare Transformation eines normalverteilten Zufallsvektors

Die Tatsache, dass ein linear transformierter normalverteilter Zufallsvektor wiederum normalverteilt ist und dass sich die Parameter der Verteilung des transformierten Zufallsvektors aus den Parametern der Verteilung des ursprünglichen Zufallsvektors sowie den Transformationsparametern bestimmen lassen, bleibt auch im Falle einer nicht notwendigerweise invertierbaren linearen Transformation und auch im Falle einer nicht notwendigerweise invertierbaren linear-affinen Transformation wahr. Dies ist die Aussage folgenden zentralen Theorems. Für einen Beweis verweisen wir auf Anderson (2003), Abschnitt 2.4.

Theorem 4.4 (Linear-affine Transformation eines normalverteilten Zufallsvektors). $\xi \sim N(\mu_\xi, \Sigma_\xi)$ sei ein normalverteilter n -dimensionaler Zufallsvektor und es sei

$$\zeta := Ax + b \text{ mit } A \in \mathbb{R}^{m \times n} \text{ und } b \in \mathbb{R}^m. \quad (4.15)$$

Dann gilt

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu + b \in \mathbb{R}^m \text{ und } \Sigma_\zeta = A\Sigma A^T \in \mathbb{R}^{m \times m}. \quad (4.16)$$

4.4 Sphärizität

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 ξ_1, \dots, ξ_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 -variaten normalverteilten Zufallsvektors ξ mit Erwartungswertparameter $\mu := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$. Dann gilt

$$p_\xi(x) = p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i) \quad (4.17)$$

und insbesondere

$$N(x; \mu, \sigma^2 I_n) = \prod_{i=1}^n N(x_i; \mu_i, \sigma^2) \quad (4.18)$$

für alle $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$.

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

$$\begin{aligned} N(x; \mu, \sigma^2 I_n) &= (2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T (\sigma^2 I_n)^{-1} (x - \mu)\right) \\ &= \left(\prod_{i=1}^n 2\pi^{-\frac{1}{2}}\right) (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^T (x - \mu)\right) \\ &= \left(\prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n N(x_i; \mu_i, \sigma^2). \end{aligned} \quad (4.19)$$

□

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 uni- und multivariaten Marginalverteilungen multivariater Normalverteilungen wiederum Normalverteilungen sind. Zum anderen lassen sich wie alle multivariaten Verteilungen multivariate Normalverteilungen 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 obige Erkenntnisse formal in den folgenden drei Theoremen zusammen.

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

$$\mu = \begin{pmatrix} \mu_v \\ \mu_\zeta \end{pmatrix} \in \mathbb{R}^n, \tag{4.20}$$

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

$$\Sigma = \begin{pmatrix} \Sigma_{vv} & \Sigma_{v\zeta} \\ \Sigma_{\zeta v} & \Sigma_{\zeta\zeta} \end{pmatrix} \in \mathbb{R}^{n \times n}, \tag{4.21}$$

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 $\zeta := (\xi_{k+1}, \dots, \xi_n)^T$ k - und l -dimensionale normalverteilte Zufallsvektoren, respektive, und es gilt

$$v \sim N(\mu_v, \Sigma_{vv}) \text{ und } \zeta \sim N(\mu_\zeta, \Sigma_{\zeta\zeta}). \tag{4.22}$$

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 z.B. Maradia, Kent, und Bibby (1979), Kapitel 3 oder Anderson (2003), Kapitel 2. Abbildung 4.6 visualisiert Theorem 4.6 für den Fall $n := 2, k := 1, l := 1$,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2 \text{ und } \Sigma := \begin{pmatrix} 0.10 & 0.08 \\ 0.08 & 0.15 \end{pmatrix} \in \mathbb{R}^{2 \times 2} \tag{4.23}$$

Abbildung 4.6 A zeigt dabei die WDF des bivariaten Zufallsvektors ξ und Abbildung 4.6 B und C die WDFen der entsprechenden marginalen Zufallsvariablen v und ζ .

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.

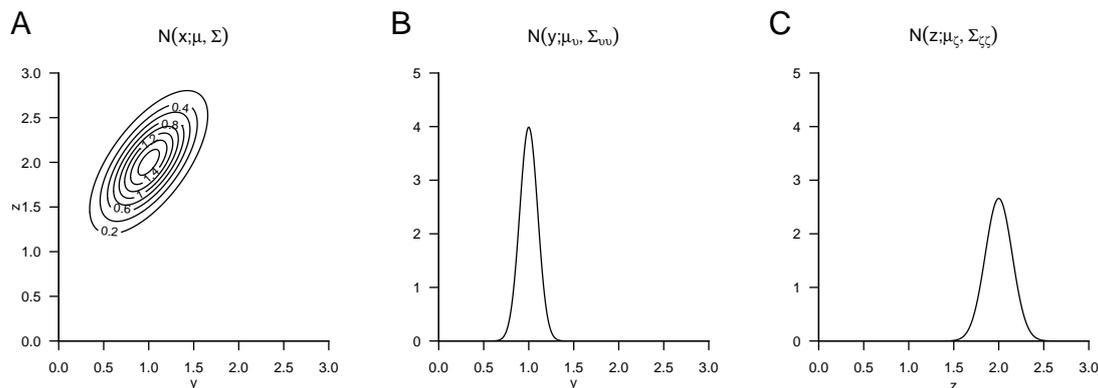


Abbildung 4.6 Marginale Verteilungen eines bivariaten normalverteilten Zufallsvektor.

Theorem 4.7 (Gemeinsame Normalverteilungen). ξ sei ein m -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_\xi : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_\xi(x) := N(x; \mu_\xi, \Sigma_{\xi\xi}) \text{ mit } \mu_\xi \in \mathbb{R}^m, \Sigma_{\xi\xi} \in \mathbb{R}^{m \times m}, \quad (4.24)$$

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

$$p_{v|\xi}(\cdot|x) : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, y \mapsto p_{v|\xi}(y|x) := N(y; A\xi + b, \Sigma_{vv}) \text{ mit } \Sigma_{vv} \in \mathbb{R}^{n \times n}. \quad (4.25)$$

Dann ist der $m + n$ -dimensionale Zufallsvektor $(\xi, v)^T$ normalverteilt mit (gemeinsamer) WDF

$$p_{\xi,v} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p_{\xi,v} \left(\begin{pmatrix} x \\ y \end{pmatrix} \right) = N \left(\begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi,v}, \Sigma_{\xi,v} \right), \quad (4.26)$$

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

$$\mu_{\xi,v} = \begin{pmatrix} \mu_\xi \\ A\mu_\xi + b \end{pmatrix} \text{ und } \Sigma_{\xi,v} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\xi}A^T \\ A\Sigma_{\xi\xi} & \Sigma_{vv} + A\Sigma_{\xi\xi}A^T \end{pmatrix}. \quad (4.27)$$

Insbesondere ergeben sich die Parameter der gemeinsamen Verteilung also als linear-affine Transformation der Parameter der induzierenden Verteilungen. Abbildung 4.7 visualisiert Theorem 4.7 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 4.7 A zeigt dabei die WDF der Zufallsvariable ξ , Abbildung 4.7 B zeigt die WDF der bedingten Verteilung der Zufallsvariable v gegeben ξ und Abbildung 4.6 C schließlich zeigt die WDFen des induzierten bivariaten Zufallsvektors (ξ, v) .

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.

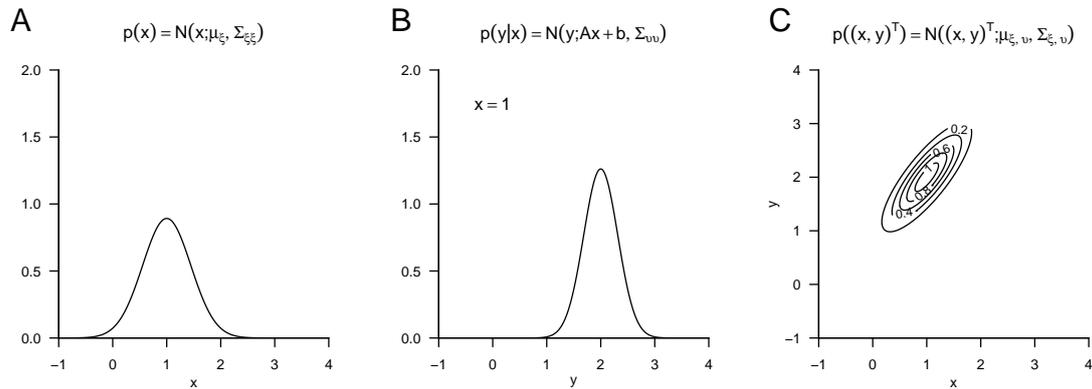


Abbildung 4.7 Gemeinsame Verteilungen einer marginalen und einer auf dieser bedingten normalverteilten Zufallsvariable.

Theorem 4.8 (Bedingte Normalverteilungen). (ξ, v) sei ein $m + n$ -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_{\xi, v} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p_{\xi, v} \left(\begin{pmatrix} x \\ y \end{pmatrix} \right) := N \left(\begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi, v}, \Sigma_{\xi, v} \right), \quad (4.28)$$

mit

$$\mu_{\xi, v} = \begin{pmatrix} \mu_{\xi} \\ \mu_v \end{pmatrix}, \Sigma_{\xi, v} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi v} \\ \Sigma_{v\xi} & \Sigma_{vv} \end{pmatrix}, \quad (4.29)$$

mit $x, \mu_{\xi} \in \mathbb{R}^m, y, \mu_v \in \mathbb{R}^n$ 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 ξ gegeben v eine m -dimensionale Normalverteilung mit bedingter WDF

$$p_{\xi|v}(\cdot|y) : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_{\xi|v}(x|y) := N(x; \mu_{\xi|v}, \Sigma_{\xi|v}) \quad (4.30)$$

mit

$$\mu_{\xi|v} = \mu_{\xi} + \Sigma_{\xi v} \Sigma_{vv}^{-1} (y - \mu_v) \in \mathbb{R}^m \quad (4.31)$$

und

$$\Sigma_{\xi|v} = \Sigma_{\xi\xi} - \Sigma_{\xi v} \Sigma_{vv}^{-1} \Sigma_{v\xi} \in \mathbb{R}^{m \times m}. \quad (4.32)$$

Im Zusammenspiel mit Theorem 4.7 und Theorem 4.6 können die Parameter bedingter und marginaler Normalverteilungen also aus den Parametern der komplementären bedingten und marginalen Normalverteilungen bestimmt werden. Abbildung 4.8 visualisiert Theorem 4.8 für den Fall $m := 2, n := 1$,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 0.12 & 0.09 \\ 0.09 & 0.12 \end{pmatrix} \quad (4.33)$$

Dabei zeigt Abbildung 4.8 A die WDF des bivariaten Zufallsvektors $(\xi, v)^T$ und Abbildung 4.8 B und C zeigen die WDF der bedingten Verteilung der Zufallsvariable ξ gegeben $v = 1.5$ und $v = 2.8$, respektive.

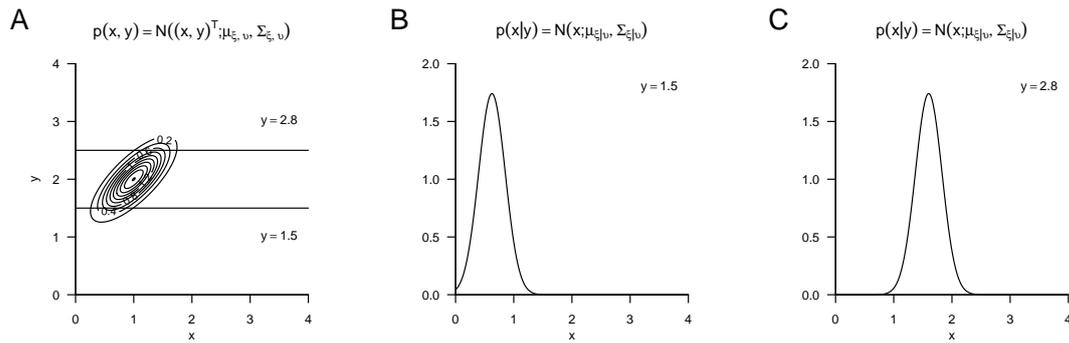


Abbildung 4.8 Bedingte Normalverteilungen

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 insbesondere auf Pearson (1896) zurück (Seal (1967)). Die ursprüngliche Formulierung der multivariaten Normalverteilung wird bei Edgeworth (1892) verortet. Tong (1990) gibt einen umfassenden Überblick zur Theorie und Anwendung der multivariaten Normalverteilung.

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. Definieren Sie die WDF eines multivariaten normalverteilten Zufallsvektors wieder und erläutern Sie diese.
5. Welche Werte haben der Erwartungswert und die Kovarianzmatrix eines normalverteilten Zufallsvektors?
6. Visualisieren Sie die WDF eines 2-dimensionalen normalverteilten Zufallsvektors mit den Parameterwerten

$$\mu := \begin{pmatrix} 10 \\ 15 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}. \quad (4.34)$$

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

$$v = X\beta + \varepsilon, \tag{5.1}$$

wobei

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

$$\varepsilon \sim N(0_n, \sigma^2 I_n). \tag{5.2}$$

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

•

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

$$y = X\beta + e \tag{5.3}$$

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

Die Gesamtzahl an Parametern des ALMs beträgt $p+1$, bestehend aus p skalaren Betaparametern und einem Varianzparameter σ^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 ε als sphärisch angenommen wird. Damit folgt direkt, dass die $\varepsilon_1, \dots, \varepsilon_n$ *unabhängige* normalverteilte Zufallsvariablen mit identischem Varianzparameter sind. Weil für ε 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 $v_i, i = 1, \dots, n$ von v nach Gleichung 5.1, dass

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

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

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

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

das ALM. Dann gilt

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

Beweis. Mit dem Theorem 4.4 gilt für

$$\varepsilon \sim N(0_n, \sigma^2 I_n) \text{ und } v := I_n \varepsilon + X\beta, \quad (5.7)$$

dass

$$v \sim N(I_n 0_n + X\beta, I_n(\sigma^2 I_n)I_n^T) = N(X\beta, \sigma^2 I_n). \quad (5.8)$$

Mit der Definition $\mu := X\beta \in \mathbb{R}^n$ folgt das Theorem dann direkt.

□

Im ALM sind die Daten v 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 v_1, \dots, v_n von v , also die Zufallsvariablen, die skalare Datenpunkte modellieren, damit unabhängige normalverteilte Zufallsvariablen der Form

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

Da im Allgemeinen aber $(X\beta)_i \neq (X\beta)_j$ für $i \neq j$ gilt, sind die $v_i, i = 1, \dots, n$ im Allgemeinen aber nicht identisch verteilt. Das Szenario unabhängig und identisch normalverteilter datenmodellierender Zufallsvariablen kann aber natürlich 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 σ^2 ,

$$v_1, \dots, v_n \sim N(\mu, \sigma^2). \quad (5.10)$$

Dann gilt, dass Gleichung 5.10 äquivalent ist zu

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

In Matrixschreibweise ist dies wiederum äquivalent zu dem ALM Spezialfall

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

In **R** können Realisierungen des ALMs leicht mithilfe eines Zufallszahlengenerators für multivariate Normalverteilungen durch Spezifikation der entsprechenden Erwartungs- und Kovarianzparameter 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.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 12 # Anzahl von Datenpunkten
4 p = 1 # Anzahl von Betaparameter
5 X = matrix(rep(1,n), nrow = n) # Designmatrix
6 I_n = diag(n) # n x n Einheitsmatrix
7 beta = 2 # wahrer, aber unbekannter, Betaparameter
8 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
9
10 # Datenrealisierung
11 y = mvrnorm(1, X %*% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs

```

Realisierungen: 0.18 3.72 3.13 -0.13 1.63 0.29 1.31 0.12 0.59 2.14 2.38 2.04

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

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

In Matrixschreibweise ist dies wiederum äquivalent zu dem ALM Spezialfall

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

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.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 10 # Anzahl von Datenpunkten
4 p = 2 # Anzahl von Betaparametern
5 x = 1:n # Prädiktorwerte
6 X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix
7 I_n = diag(n) # n x n Einheitsmatrix
8 beta = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter, Betaparameter
9 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
10
11 # Datenrealisierung
12 y = mvrnorm(1, X %*% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs

```

Wir visualisieren obige Realisierung in Abbildung 5.1.

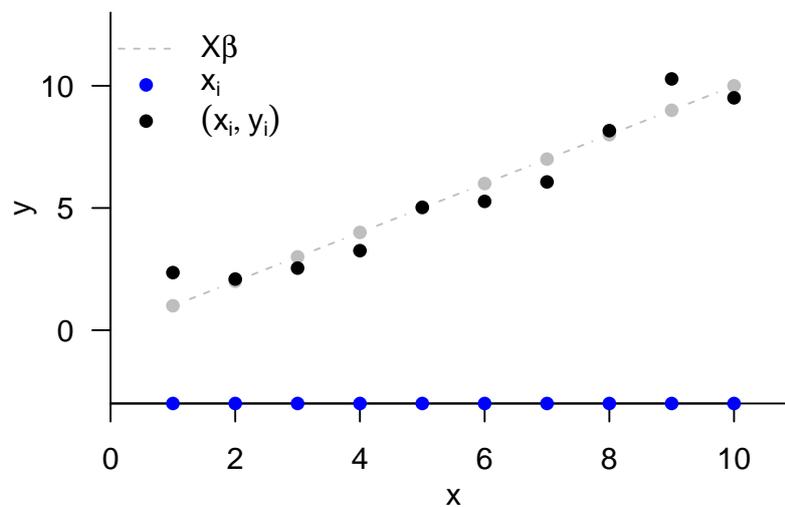


Abbildung 5.1 Realisierung einer einfachen linearen Regression.

5.2 Identifizierbarkeit und Schätzbarkeit

Wir haben oben gesehen, dass die Datenverteilung des ALM durch

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

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 μ .

•

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 μ . β 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 Designmatrixrang). $v \sim N(X\beta, \sigma^2 I_n)$ sei die Datenverteilung eines ALMs mit $\text{rg}(X) = p$. Dann ist β 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 μ . 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 β , 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 Beta-parametrisierung von μ . 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 ϕ existiert, so dass

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

Beweis. Wir zeigen die Aussage lediglich in eine Richtung.

\Rightarrow Wir nehmen an, es existiert eine Funktion ϕ , 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 β , 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, so dass

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

•

Diese Definition erschließt sich wie folgt: Nach dem Theorem zu identifizierbaren Funktionen muss eine identifizierbare lineare Funktion von β 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 ϕ eine lineare Funktion sein. Damit kann aber auch ϕ 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 öffnet 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.

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

also

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

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

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

also

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

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 repräsentieren. 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 –1en enthält. In diesem Fall repräsentieren die Betaparameter Gruppenerwartungswerte und wir werden sehen, dass die Betaparameterschätzer in der Repräsentationen von Gruppenstichprobenmitteln resultieren. 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* 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). Seal (1967) gibt einen ausführlichen Überblick zur Geschichte des Allgemeinen Linearen Modells im 19. und 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. Geben Sie die Definition des Allgemeinen Linearen Modells wieder.
2. Erläutern Sie die deterministischen und probabilistischen Aspekte des ALMs.
3. Wie viele skalare Parameter hat das ALM mit sphärischer Kovarianzmatrix?
4. Warum sind die Komponenten des ALM Zufallsfehlers unabhängig und identisch verteilt?
5. Geben Sie das Theorem zur Datenverteilung des Allgemeinen Linearen Modells wieder.
6. Sind die Komponenten des ALM Datenvektors immer unabhängig und identisch verteilt?
7. Schreiben Sie das Szenario von n unabhängig und identisch normalverteilten Zufallsvariablen in ALM Form.
8. 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 Parameterschätzerverteilungen des ALMs.

6.1 Betaparameterschätzung

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

Theorem 6.1 (Betaparameterschätzer). *Es sei*

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

das Allgemeine Lineare Modell und es sei

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

Dann gelten

(1) $\hat{\beta}$ ist der KQ-Schätzer von $\beta \in \mathbb{R}^p$, es gilt also

$$\hat{\beta} = \operatorname{argmin}_{\tilde{\beta}} (v - X\tilde{\beta})^T (v - X\tilde{\beta}). \quad (6.3)$$

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

Beweis. (1) Wir zeigen in einem ersten Schritt, dass $\hat{\beta}$ ein KQ-Schätzer ist, dass also $\hat{\beta}$ die Summe der Abweichungsquadrate

$$(v - X\tilde{\beta})^T (v - X\tilde{\beta}) = \sum_{i=1}^n (v_i - (X\tilde{\beta})_i)^2 \quad (6.4)$$

minimiert (die Notation $\tilde{\beta}$ für das Minimierungsargument dient hier lediglich dazu, es vom wahren, aber unbekanntem, Parameterwert $\beta \in \mathbb{R}^p$ abzugrenzen und ist ansonsten ohne Bedeutung). Dazu halten wir zunächst fest, dass

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

Weiterhin gilt dann auch, dass

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

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

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

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

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

Es ergibt sich

$$\begin{aligned} & (v - X\tilde{\beta})^T(v - X\tilde{\beta}) \\ &= (v - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta})^T(v - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta}) \\ &= ((v - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta}))^T((v - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta})) \\ &= (v - X\hat{\beta})^T(v - X\hat{\beta}) + (v - X\hat{\beta})^T X(\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T(v - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\ &= (v - X\hat{\beta})^T(v - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \\ &= (v - X\hat{\beta})^T(v - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \end{aligned} \quad (6.9)$$

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

$$(\hat{\beta} - \tilde{\beta})^T X^T X(\hat{\beta} - \tilde{\beta}) \geq 0 \quad (6.10)$$

nimmt dieser Term genau dann seinen Minimalwert 0 an, wenn

$$(\hat{\beta} - \tilde{\beta}) = 0_p \Leftrightarrow \tilde{\beta} = \hat{\beta}. \quad (6.11)$$

Also gilt

$$\hat{\beta} = \operatorname{argmin}_{\tilde{\beta}} (v - X\tilde{\beta})^T(v - X\tilde{\beta}). \quad (6.12)$$

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

$$\ell : \mathbb{R}^p \rightarrow \mathbb{R}, \tilde{\beta} \mapsto \ln p_{\tilde{\beta}}(v) = \ln N(y; X\tilde{\beta}, \sigma^2 I_n) \quad (6.13)$$

wobei gilt, dass

$$\begin{aligned} \ln N(y; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left((2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \right) \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta}). \end{aligned} \quad (6.14)$$

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

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}((X^T X)^{-1} X^T v) = (X^T X)^{-1} X^T \mathbb{E}(v) = (X^T X)^{-1} X^T X \beta = \beta. \quad (6.15)$$

□

Theorem 6.1 gibt mit

$$\hat{\beta} = (X^T X)^{-1} X^T v \quad (6.16)$$

eine Formel an, um β anhand der Designmatrix und einer Realisierung $y \in \mathbb{R}^n$ von v konkret zu schätzen. $\hat{\beta}$ ist ein unverzerrter Schätzer von β und als Maximum-Likelihood Schätzer insbesondere auch konsistent, asymptotisch normalverteilt und asymptotisch effizient. Wir sehen an späterer Stelle dass $\hat{\beta}$ sogar normalverteilt ist. Neben den genannten Eigenschaften hat $\hat{\beta}$ noch weitere gute Eigenschaften. Zum Beispiel besitzt $\hat{\beta}$ die kleinste Varianz in der Klasse der linearen unverzerrten Schätzer von β . 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). Es sei

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.17)$$

das Allgemeine Lineare Modell und es sei

$$\hat{\beta} := (X^T X)^{-1} X^T v. \quad (6.18)$$

der Betaparameterschätzer. Dann heißt der Zufallsvektor

$$\hat{v} := X\hat{\beta} = X(X^T X)^{-1} X^T v \quad (6.19)$$

die *erklärten Daten*, der Zufallsvektor

$$\hat{\varepsilon} := v - \hat{v} = v - X\hat{\beta} \quad (6.20)$$

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

$$\hat{\varepsilon}_i := v_i - \hat{v}_i = v_i - (X\hat{\beta})_i \quad (6.21)$$

die *Residuen*.

•

6.2 Varianzparameterschätzung

Wir fassen die Frequentistische Punktschätzung des Varianzparameters in folgendem Theorem zusammen, das wir nicht beweisen wollen.

Theorem 6.2. *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.22)$$

das Allgemeine Lineare Modell. Dann ist

$$\hat{\sigma}^2 := \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - p} \quad (6.23)$$

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

Theorem 6.2 gibt mit

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - p} = \frac{(v - X\hat{\beta})^T (v - X\hat{\beta})}{n - p}. \quad (6.24)$$

eine Formel an, um σ^2 anhand der Designmatrix, des Betaparameterschätzers und einer Realisierung von v zu schätzen. Offenbar gilt mit Theorem 6.2, dass

$$\hat{\sigma}^2 = \frac{1}{n - p} \sum_{i=1}^n (v_i - (X\hat{\beta})_i)^2. \quad (6.25)$$

$\hat{\sigma}^2$ wird also durch die Summe der quadrierten Residuen, also als eine Summe von Abweichungsquadraten geschätzt. Für einen Beweis von Theorem 6.2 verweisen wir z.B. auf S. R. Searle (1971), S. R. Searle und Gruber (2017) oder Rencher und Schaalje (2008). Aus probabilistischer Perspektive handelt es sich bei $\hat{\sigma}^2$ nicht um einen Maximum-Likelihood Schätzer, sondern um einen *Restricted Maximum-Likelihood Schätzer* von σ^2 (vgl. Harville (1977), Foulley (1993), Starke und Ostwald (2017)). Aus geometrischer Perspektive handelt es sich bei $\hat{\sigma}^2$ um einen KQ-Schätzer (vgl. Christensen (2011), Kapitel 2).

6.3 Unabhängig identisch normalverteilte Zufallsvariablen

Als erste Anwendung von Theorem 6.1 und Theorem 6.2 analysieren wir das Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2 ,

$$v_i \sim N(\mu, \sigma^2) \text{ für } i = 1, \dots, n. \quad (6.26)$$

Schreibt man dieses Modell in seiner Designmatrixform (vgl. Gleichung 5.10) dann gilt, wie unten gezeigt,

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n v_i =: \bar{v} \text{ und } \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2 =: s_v^2 \quad (6.27)$$

In diesem Fall ist also der Betaparameterschätzer mit dem Stichprobenmittel \bar{v} der v_1, \dots, v_n und der Varianzparameterschätzer mit der Stichprobenvarianz s_v^2 der v_1, \dots, v_n identisch.

Gleichung 6.27 ergibt sich wie folgt. Zum einen gilt

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T v \\ &= (1_n^T 1_n)^{-1} 1_n^T v \\ &= \left((1 \quad \dots \quad 1) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)^{-1} (1 \quad \dots \quad 1) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ &= n^{-1} \sum_{i=1}^n v_i \\ &= \frac{1}{n} \sum_{i=1}^n v_i \\ &=: \bar{v}. \end{aligned} \quad (6.28)$$

Zum anderen gilt

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-1} (v - X\hat{\beta})^T (v - X\hat{\beta}) \\ &= \frac{1}{n-1} (v - 1_n \bar{v})^T (v - 1_n \bar{v}) \\ &= \frac{1}{n-1} \left(\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{v} \right)^T \left(\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{v} \right) \\ &= \frac{1}{n-1} (v_1 - \bar{v} \quad \dots \quad v_n - \bar{v}) \begin{pmatrix} v_1 - \bar{v} \\ \vdots \\ v_n - \bar{v} \end{pmatrix} \\ &= \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2 \\ &=: s_v^2. \end{aligned} \quad (6.29)$$

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

```

1 # Modellformulierung
2 library(MASS)
3 n = 12
4 p = 1
5 X = matrix(rep(1,n), nrow = n)
6 I_n = diag(n)
7 beta = 2
# Multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter, Betaparameter

```

```

8  sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
9
10 # Datenrealisierung
11 y = mvrnorm(1, X %>% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs
12
13 # Modellschätzung
14 beta_hat = solve(t(X) %>% X) %>% t(X) %>% y # Betaparameterschätzer
15 eps_hat = y - X %>% beta_hat # Residuenvektor
16 sigsq_hat = (t(eps_hat) %>% eps_hat) / (n-p) # Varianzparameterschätzer

beta : 2
hat{beta} : 1.347296
sigsqr : 1
hat{sigsqr}: 0.7921628

```

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

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 12 # Anzahl Datenpunkte
4 p = 1 # Anzahl Betaparameter
5 X = matrix(rep(1,n), nrow = n) # Designmatrix
6 I_n = diag(n) # n x n Einheitsmatrix
7 beta = 2 # wahrer, aber unbekannter, Betaparameter
8 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
9
10 # Frequentistische Simulation
11 nsim = 1e4 # Anzahl Datenrealisierungen
12 beta_hat = rep(NA, nsim) # \hat{\beta} Realisierungsarray
13 sigsq_hat = rep(NA, nsim) # \hat{sigsqr} Realisierungsarray
14 for(i in 1:nsim){
15   y = mvrnorm(1, X * beta, sigsq*I_n) # Datenrealisierung
16   beta_hat[i] = solve(t(X) %>% X) %>% t(X) %>% y # Betaparameterschätzer
17   eps_hat = y - X %>% beta_hat[i] # Residuenvektor
18   sigsq_hat[i] = (t(eps_hat) %>% eps_hat) / (n-p) # Varianzparameterschätzer
19 }

Wahrer, aber unbekannter, Betaparameter : 2
Geschätzter Erwartungswert des Betaparameterschätzers : 1.998824
Wahrer, aber unbekannter, Varianzparameter : 1
Geschätzter Erwartungswert des Varianzparameterschätzers : 0.9995151

```

6.4 Einfache lineare Regression

Als zweite Anwendung von Theorem 6.1 und Theorem 6.2 analysieren wir das Szenario der einfachen linearen Regression

$$v_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n. \quad (6.30)$$

Basierend auf der Designmatrixform (vgl. Gleichung 5.15) dieses Modells ergibt sich, wie unten gezeigt,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{v} - \frac{c_{xv}}{s_x^2} \bar{x} \\ \frac{c_{xv}}{s_x^2} \end{pmatrix} \text{ und } \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (v_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \quad (6.31)$$

wobei \bar{x} und \bar{v} die Stichprobenmittel der x_1, \dots, x_n und v_1, \dots, v_n , c_{xv} die Stichprobenkovarianz der x_1, \dots, x_n und v_1, \dots, v_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 respektiven 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 den Wert

$$q(\tilde{\beta}) = \sum_{i=1}^n (v_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2 = (v - X\tilde{\beta})^T (v - X\tilde{\beta}) \quad (6.32)$$

hinsichtlich $\tilde{\beta}$ minimieren.

Um die Form des Betaparameterschätzers in Gleichung 6.31 herzuleiten, halten wir zunächst fest, dass

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(v_i - \bar{v}) &= \sum_{i=1}^n (x_i v_i - x_i \bar{v} - \bar{x} v_i + \bar{x} \bar{v}) \\ &= \sum_{i=1}^n x_i v_i - \sum_{i=1}^n x_i \bar{v} - \sum_{i=1}^n \bar{x} v_i + \sum_{i=1}^n \bar{x} \bar{v} \\ &= \sum_{i=1}^n x_i v_i - \bar{v} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n v_i + n \bar{x} \bar{v} \\ &= \sum_{i=1}^n x_i v_i - \bar{v} n \bar{x} - \bar{x} n \bar{v} + n \bar{x} \bar{v} \\ &= \sum_{i=1}^n x_i v_i - n \bar{x} \bar{v} - n \bar{x} \bar{v} + n \bar{x} \bar{v} \\ &= \sum_{i=1}^n x_i v_i - n \bar{x} \bar{v}, \end{aligned} \quad (6.33)$$

Weiterhin halten wir fest, dass

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2x_i \bar{x} + \sum_{i=1}^n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} n \bar{x} + n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n \bar{x}^2 + n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n \bar{x}^2. \end{aligned} \quad (6.34)$$

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

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T v \\ &= \left(\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ &= \left(\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \right)^{-1} \begin{pmatrix} \sum_{i=1}^n v_i \\ \sum_{i=1}^n x_i v_i \end{pmatrix} \\ &= \begin{pmatrix} n & n \bar{x} \\ n \bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n \bar{v} \\ \sum_{i=1}^n x_i v_i \end{pmatrix}. \end{aligned} \quad (6.35)$$

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

$$\frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad (6.36)$$

weil

$$\begin{aligned} & \frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} \frac{n s_x^2}{n} + n\bar{x}^2 - n\bar{x}^2 & \frac{s_x^2 n\bar{x}}{n} + n\bar{x}^2 \bar{x} - \bar{x} \sum_{i=1}^n x_i^2 \\ -\bar{x}n + n\bar{x} & -n\bar{x}^2 + \sum_{i=1}^n x_i^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & s_x^2 \bar{x} - \bar{x} (\sum_{i=1}^n x_i^2 - n\bar{x}^2) \\ 0 & \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & s_x^2 \bar{x} - \bar{x} s_x^2 \\ 0 & s_x^2 \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} s_x^2 & 0 \\ 0 & s_x^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \quad (6.37)$$

Es ergibt sich also

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_x^2} & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix} \begin{pmatrix} n\bar{v} \\ \sum_{i=1}^n x_i v_i \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2} \right) n\bar{v} - \frac{\bar{x} \sum_{i=1}^n x_i v_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i v_i}{s_x^2} - \frac{n\bar{x}\bar{v}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{n\bar{v}}{n} + \frac{\bar{x}^2 n\bar{v}}{s_x^2} - \frac{\bar{x} \sum_{i=1}^n x_i v_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i v_i - n\bar{x}\bar{v}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{v} + \frac{\bar{x} n\bar{x}\bar{v} - \bar{x} \sum_{i=1}^n x_i v_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i v_i - n\bar{x}\bar{v}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{v} - \frac{\sum_{i=1}^n x_i v_i - n\bar{x}\bar{v}}{s_x^2} \bar{x} \\ \frac{\sum_{i=1}^n x_i v_i - n\bar{x}\bar{v}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{v} - \frac{c_{xv}}{s_x^2} \bar{x} \\ \frac{c_{xv}}{s_x^2} \end{pmatrix}. \end{aligned} \quad (6.38)$$

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.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 10 # Anzahl Datenpunkte
4 p = 2 # Anzahl Betaparameter
5 x = 1:n # Prädiktorwerte
6 X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix
7 I_n = diag(n) # n x n Einheitsmatrix
8 beta = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter, Betaparameter
9 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
10
11 # Datenrealisierung
12 y = mvrnorm(1, X %*% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs
13
14 # Parameterschätzung
15 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
16 eps_hat = y - X %*% beta_hat # Residuenvektor
17 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer

beta : 0 1
hat{beta} : 0.7426838 0.9830537
sigsqr : 1
hat{sigsqr}: 0.9961321

```

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

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 10 # Anzahl Datenpunkte
4 p = 2 # Anzahl Betaparameter
5 x = 1:n # Prädiktorwerte
6 X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix
7 I_n = diag(n) # n x n Einheitsmatrix
8 beta = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter, Betaparameter
9 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
10
11 # Frequentistische Simulation
12 nsim = 1e4 # Anzahl Realisierungen des n-dimensionalen ZVs
13 beta_hat = matrix(rep(NA,n*p*nsim), nrow = p) # \hat{\beta} Realisierungsarray
14 sigsq_hat = rep(NA,nsim) # \hat{\sigma^2} Realisierungsarray
15 for(i in 1:nsim){ # Simulationsiterationen
16 y = mvrnorm(1, X %*% beta, sigsq*I_n) # Datenrealisierung
17 beta_hat[,i] = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
18 eps_hat = y - X %*% beta_hat[,i] # Residuenvektor
19 sigsq_hat[i] = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
20 }

```

Wahrer, aber unbekannter, Betaparameter : 0 1
Geschätzter Erwartungswert des Betaparameterschätzers : -0.00191267 0.9997315
Wahrer, aber unbekannter, Varianzparameter : 1
Geschätzter Erwartungswert des Varianzparameterschätzers : 1.000095

6.5 Frequentistische Schätzerverteilungen

Wir dokumentieren die Frequentistische Verteilung des Betaparameterschätzers in folgendem Theorem.

Theorem 6.3 (Frequentistische Verteilung des Betaparameterschätzers). *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.39)$$

das ALM. Weiterhin sei

$$\hat{\beta} := (X^T X)^{-1} X^T v \quad (6.40)$$

der Betaparameterschätzer. Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (6.41)$$

Beweis. Das Theorem folgt direkt mit dem Theorem zur linearen Transformation von multivariaten Normalverteilungen. Speziell gilt hier:

$$\hat{\beta} \sim N((X^T X)^{-1} X^T X \beta, (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T). \quad (6.42)$$

Der Erwartungswertparameter vereinfacht sich dann zu

$$(X^T X)^{-1} X^T X \beta = \beta. \quad (6.43)$$

Der Kovarianzmatrixparameter vereinfacht sich wie folgt.

$$\begin{aligned} (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T &= (X^T X)^{-1} X^T (\sigma^2 I_n) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned} \quad (6.44)$$

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

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (6.45)$$

□

Mit Theorem 6.3 folgt also insbesondere auch für den Erwartungswert und die Kovarianzmatrix des Betaparameterschätzers, dass

$$\mathbb{E}(\hat{\beta}) = \beta \text{ und } \mathbb{C}(\hat{\beta}) = \sigma^2(X^T X)^{-1}. \quad (6.46)$$

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, welches wir nicht beweisen wollen.

Theorem 6.4 (Frequentistische Verteilung des Varianzparameterschätzers). *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (6.47)$$

das ALM. Weiterhin sei

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - p} \quad (6.48)$$

der Varianzparameterschätzer. Dann gilt

$$\frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n - p). \quad (6.49)$$

Da es sich bei $(n - p)\hat{\sigma}^2$ um eine Summe normalverteilter Zufallsvariablen handelt, liegt die χ^2 -Verteilung im Lichte der χ^2 -Transformation bei normalverteilten Zufallsvariablen zumindest nahe. Allerdings ist $\hat{\sigma}^2$ selbst nicht χ^2 verteilt, sondern lediglich seine durch Multiplikation mit $\frac{n-p}{\sigma^2}$ skalierte Version. Wir wollen die Frequentistischen Schätzerverteilungen aus Theorem 6.3 und Theorem 6.4 noch an den beiden Standardbeispielen verdeutlichen.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$v \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \quad (6.50)$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen bei bekannter Varianz. Wir haben bereits gesehen, dass in diesem Fall $\hat{\beta}$ mit dem Stichprobenmittel \bar{v} identisch ist. Theorem 6.3 impliziert dann mit

$$(X^T X)^{-1} = (\mathbf{1}_n^T \mathbf{1}_n)^{-1} = \frac{1}{n} \quad (6.51)$$

dass

$$\bar{v} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (6.52)$$

Das Stichprobenmittel von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2 ist also normalverteilt mit Erwartungswertparameter μ und Varianzparameter σ^2/n . Wir haben diese Tatsache bereits im Kontext der Transformationen der Normalverteilungen unter dem Begriff der *Mittelwertstransformation* gesehen.

Beispiel (2) Einfache lineare Regression

Es sei

$$v \sim N(X\beta, \sigma^2 I_n) \text{ mit } \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2 \text{ und } \sigma^2 > 0 \quad (6.53)$$

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

$$\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{s_{xx}} \begin{pmatrix} \frac{s_{xx}}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \text{ mit } s_{xx} := \sum_{i=1}^n (x_i - \bar{x})^2. \quad (6.54)$$

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. Folgender **R** Code simuliert die frequentistischen Verteilungen von Beta- und Varianzparameterschätzern im Szenario der einfachen linearen Regression.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 10 # Anzahl von Datenpunkten
4 p = 2 # Anzahl von Betparametern
5 x = 1:n # Prädiktorwerte
6 X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix
7 I_n = diag(n) # n x n Einheitsmatrix
8 beta = matrix(c(0,1), nrow = p) # wahrer, aber unbekannter, Betaparameter
9 sigsq = .5 # wahrer, aber unbekannter, Varianzparameter
10
11 # Frequentistische Simulation
12 nsim = 10 # Anzahl Realisierungen n-dimensionaler ZV
13 y = matrix(rep(NaN,n*nsim), nrow = n) # y Realisierungsarray
14 beta_hat = matrix(rep(NaN,p*nsim), nrow = p) # \hat{\beta} Realisierungsarray
15 for(i in 1:nsim){
16   y[,i] = mvrnorm(1,X %*% beta, sigsq*I_n) # eine Realisierung n-dimensionaler ZV
17   beta_hat[,i] = solve(t(X)%*%X)%*%t(X)%*%y[,i] # \hat{\beta} = (X^T X)^{-1} X^T y
18 }

```

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 Varianzparameter-schätzung ALM 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. Starke und Ostwald (2017), Lindholm und Wahl (2020)).

6.7 Selbstkontrollfragen

1. Geben Sie das Theorem zum Betaparameterschätzer wieder.
2. Warum ist der Betaparameterschätzer ein Maximum-Likelihood Schätzer?

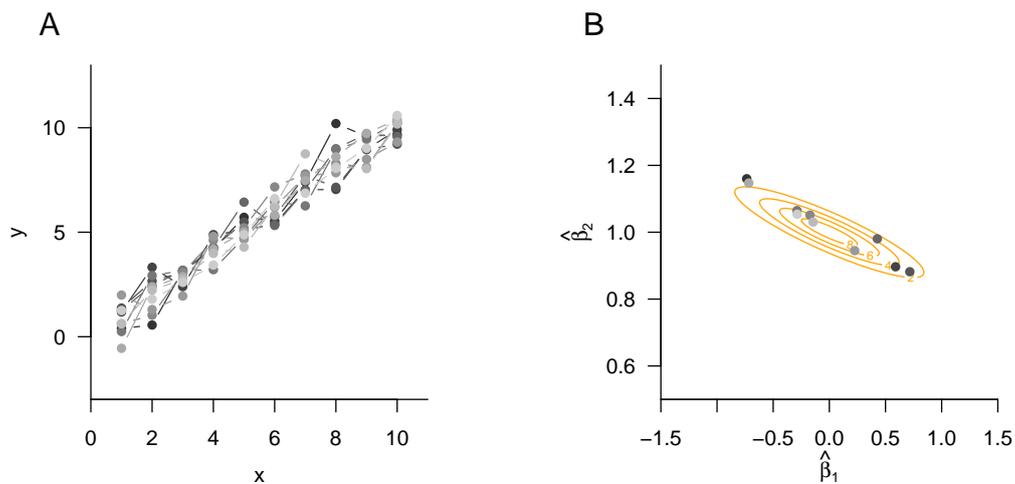


Abbildung 6.1 Frequentistische Betaparameterschätzerverteilung bei einfacher linearer Regression

3. Geben Sie das Theorem zum Varianzparameterschätzer wieder.
4. Geben Sie die Parameterschätzer bei n unabhängigen und identisch normalverteilten Zufallsvariablen an.
5. Geben Sie die Parameterschätzer bei einfacher linearer Regression an.
6. Geben Sie das Theorem zur Verteilung des Betaparameterschätzers wieder.
7. 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 quantifizieren dabei die geschätzten Effekte des Betaparameterschätzers in Bezug 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 Einstich- und Zweistichproben T-Tests ist 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). Es sei

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7.1)$$

das ALM. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T v \text{ und } \hat{\sigma}^2 := \frac{(v - X\hat{\beta})^T (v - X\hat{\beta})}{n - p} \quad (7.2)$$

die Betaparameter- und Varianzparameterschätzer, respektive. Dann ist für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ und einen Parameter $\beta_0 \in \mathbb{R}^p$ 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 Kontrastgewichtvektor. Offenbar dient der Kontrastgewichtvektor dazu, den Zufallsvektor $\hat{\beta} \in \mathbb{R}^p$ in die Zufallsvariable $c^T \hat{\beta}$ zu transformieren und sichert damit die Skalarität der T-Statistik. Weiterhin erlaubt die Wahl von p -dimensionalen Einheitsvektoren für den Kontrastgewichtvektor 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. Beispielfhaft 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 (vgl. Theorem 6.3).

Wählt man für β_0 dagegen β , 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 β_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. Diese ist der zentrale Inhalt folgenden Theorems, auf dessen Beweis wir verzichten.

Theorem 7.1 (T-Statistik). *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7.5)$$

das ALM. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T v \text{ und } \hat{\sigma}^2 := \frac{(v - X\hat{\beta})^T (v - X\hat{\beta})}{n - p} \quad (7.6)$$

die Betaparameter- und Varianzparameterschätzer, respektive. Schließlich sei für einen Kontrastgewichtsvektor $c \in \mathbb{R}^p$ und einen Parameter $\beta_0 \in \mathbb{R}^p$

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \quad (7.7)$$

die T-Statistik. Dann gilt

$$T \sim t(\delta, n - p) \text{ mit } \delta := \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}}. \quad (7.8)$$

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 TestszENARIO 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 TestszENARIO 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 des 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

$$v \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \quad (7.9)$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen und es seien $c := \mathbf{1}$ und $\beta_0 := \mu_0 \in \mathbb{R}$. Dann gilt für die T-Statistik

$$T = \frac{c^T \hat{\beta} - c^T \mu_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{\mathbf{1}^T \bar{v} - \mathbf{1}^T \mu_0}{\sqrt{s_v^2 \mathbf{1}^T (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}}} = \sqrt{n} \left(\frac{\bar{v} - \mu_0}{s_v} \right). \quad (7.10)$$

Dies entspricht offenbar der bekannten Einstichproben-T-Teststatistik. 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. Folgender **R** Code simuliert die Frequentistische Verteilung dieser T-Statistik für die Fälle $\beta = \beta_0$ und $\beta \neq \beta_0$.

```

1 # Modellformulierung
2 library(MASS) # multivariate Normalverteilung
3 n = 12 # Anzahl von Datenpunkten
4 p = 1 # Anzahl von Betaparametern
5 X = matrix(c(rep(1,n)), nrow = n) # Designmatrix
6 I_n = diag(n) # Einheitsmatrix
7 beta = c(0,1) # wahre, aber unbekannte, Betaparameter
8 nscn = length(beta) # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
9 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
10 c = 1 # Kontrastvektor von Interesse
11 beta_0 = 0 # Nullhypotesenbetaparameter
12
13 # Frequentistische Simulation
14 nsim = 1e4 # Anzahl Simulationen
15 delta = rep(NA, nscn) # Anzahl Nichtzentralitätsparameter
16 Tee = matrix(rep(NA, nscn*nsim), ncol = nscn) # T-Teststatistik Realisierungsarray
17 for(s in 1:nscn){ # Hypothesenszenarien
18   delta[s] = ((t(c) %*% beta[s] - t(c) %*% beta_0) / # Nichtzentralitätsparameter
19     sqrt(sigsq*t(c)%*%solve(t(X)%*%X)%*%c))
20   for(i in 1:nsim){ # Simulationsiterationen
21     y = mvrnorm(1, X %*% beta[s], sigsq*I_n) # y
22     beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # \hat{\beta}
23     eps_hat = y - X %*% beta_hat # \hat{\epsilon}
24     sigsq_hat = (t(eps_hat) %*% eps_hat)/(n-p) # \hat{\sigma}^2
25     Tee[i,s] = ((t(c) %*% beta_hat - t(c) %*% beta_0) / # T
26       sqrt(sigsq_hat*t(c)%*%solve(t(X)%*%X)%*%c))
27   }
28 }

```

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

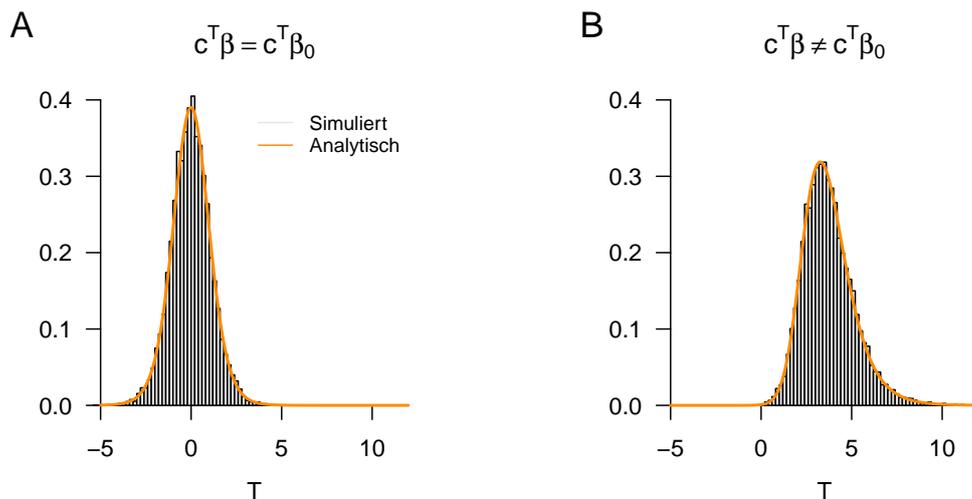


Abbildung 7.1 Verteilungen der T-Statistik bei unabhängig und identisch normalverteilten Zufallsvariablen

darstellt. Dazu betrachten wir das bekannte Beispielmodell der einfachen linearen Regression (vgl. Kapitel 5), in diesem Fall mit den wahren, aber unbekanntem, Parameterwerten $\beta_A := (1, 0)$ und $\beta_B := (1, 1)$. Weiterhin betrachten wir den Kontrastgewichtsvektor $c := (0, 1)$, so dass die T-Statistik zur 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 β_A gilt, dass $c^T \beta = c^T \beta_0$ und im Fall von β_B gilt, dass $c^T \beta \neq c^T \beta_0$. Folgender **R** Code implementiert die skizzierten Szenarien, Abbildung 7.2 A und B zeigen die resultierenden simulierten und analytischen Verteilungen der T-Statistik.

```

1 # Modellformulierung
2 library(MASS)
3 n = 10
4 p = 2
5 x = 1:n
6 X = matrix(c(rep(1,n),x), ncol = p)
7 I_n = diag(n)
8 beta = matrix(c(1,0,1,1), nrow = 2)
9 nscn = ncol(beta)
10 sigsq = 1
11 c = matrix(c(0,1), nrow = 2)
12 beta_0 = matrix(c(0,0), nrow = 2)
13
14 # Frequentistische Simulation
15 nsim = 1e4
16 delta = rep(NA, nscn)
17 Tee = matrix(rep(NA, nscn*nsim), ncol = nscn)
18 for(s in 1:nscn){
19   delta[s] = ((t(c) %>% beta[,s] - t(c) %>% beta_0) /
20             sqrt(sigsqr*t(c) %>% solve(t(X) %>% X) %>% c))
21   for(i in 1:nsim){
22     y = mvrnorm(1, X %>% beta[,s], sigsq*I_n)
23     beta_hat = solve(t(X) %>% X) %>% t(X) %>% y
24     eps_hat = y - X %>% beta_hat
25     sigsq_hat = (t(eps_hat) %>% eps_hat) / (n-p)
26     Tee[i,s] = ((t(c) %>% beta_hat - t(c) %>% beta_0) /
27               sqrt(sigsqr_hat*t(c) %>% solve(t(X) %>% X) %>% c))
28   }
29 }

```

```

# multivariate Normalverteilung
# Anzahl von Datenpunkten
# Anzahl von Betaparametern
# Prädiktorwerte
# Designmatrix
# Einheitsmatrix
# wahre , aber unbekannte , Betaparameter
# Anzahl wahrer, aber unbekannter, Hypothesenszenarien
# wahrer, aber unbekannter, Varianzparameter
# Kontrastvektor von Interesse
# Nullhypothesebetaparameter

# Anzahl Simulationen
# Anzahl Nichtzentralitätsparameter
# T-Teststatistik Realisierungsarray
# Hypothesenszenarien
# Nichtzentralitätsparameter

# Simulationsiterationen
# y
# \hat{\beta}
# \hat{\epsilon}
# \hat{\sigma}^2
# T

```

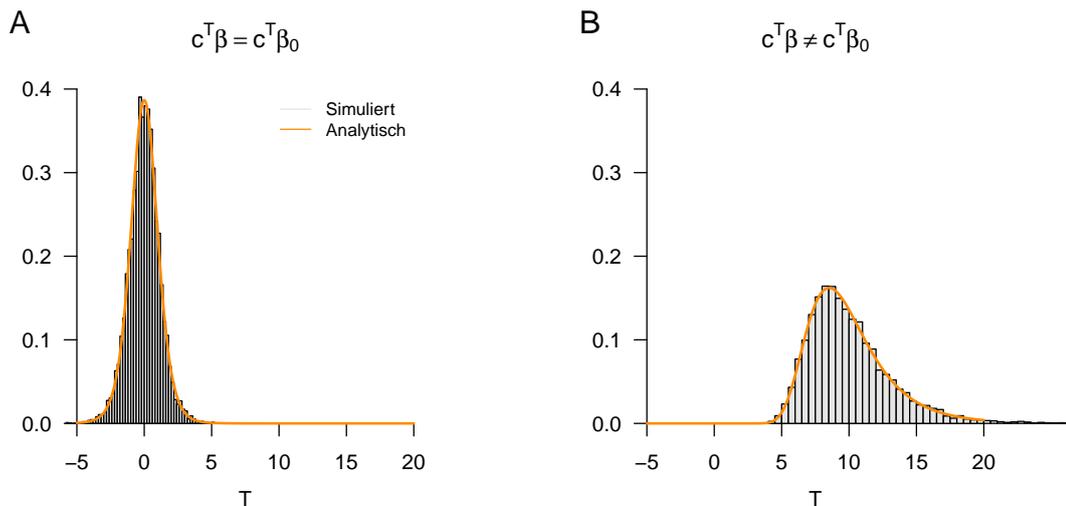


Abbildung 7.2 Verteilungen der T-Statistik bei einfacher linearer Regression

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). *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (7.11)$$

das ALM,

$$\hat{\beta} := (X^T X)^{-1} X^T v \text{ und } \hat{\sigma}^2 := \frac{(v - X\hat{\beta})^T (v - X\hat{\beta})}{n - p} \quad (7.12)$$

seien die Betaparameter- und Varianzparameterschätzer, respektive und für ein $\delta \in]0, 1[$ sei

$$t_\delta := \Psi^{-1} \left(\frac{1 + \delta}{2}; n - p \right). \quad (7.13)$$

Schließlich sei für $j = 1, \dots, p$

$$\lambda_j := \left((X^T X)^{-1} \right)_{jj} \text{ das } j\text{te Diagonalelement von } (X^T X)^{-1}. \quad (7.14)$$

Dann ist für $j = 1, \dots, p$

$$\kappa_j := \left[\hat{\beta}_j - \hat{\sigma} \sqrt{\lambda_j} t_\delta, \hat{\beta}_j + \hat{\sigma} \sqrt{\lambda_j} t_\delta \right] \quad (7.15)$$

ein δ -Konfidenzintervall für die j te Komponente β_j des Betaparameters $\beta = (\beta_1, \dots, \beta_p)^T$.

Beweis. Wir müssen zeigen, dass

$$\mathbb{P}(\kappa_j \ni \beta_j) = \delta. \quad (7.16)$$

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-Teststatistik für $T \sim t(\delta, n - p)$ gilt, dass

$$T = \frac{e_j^T \hat{\beta} - e_j^T \beta}{\sqrt{\hat{\sigma}^2 e_j^T (X^T X)^{-1} e_j}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 ((X^T X)^{-1})_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{\lambda_j}} =: T_j. \quad (7.17)$$

und

$$\delta = \frac{e_j^T \beta - e_j^T \beta}{\sqrt{\hat{\sigma}^2 e_j^T (X^T X)^{-1} e_j}} = 0 \quad (7.18)$$

Damit gilt dann auch sofort, dass $T_j \sim t(n - p)$. Weiterhin erinnern wir daran (vgl. (11) Konfidenzintervalle in Wahrscheinlichkeitstheorie und Frequentistischer Inferenz), dass per Definition von t_δ gilt, dass

$$\mathbb{P}(-t_\delta \leq T_j \leq t_\delta) \quad (7.19)$$

Aus der Definition eines δ -Konfidenzintervalls folgt dann

$$\begin{aligned} \delta &= \mathbb{P}(-t_\delta \leq T_j \leq t_\delta) \\ &= \mathbb{P}\left(-t_\delta \leq \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{\lambda_j}} \leq t_\delta\right) \\ &= \mathbb{P}\left(-t_\delta \hat{\sigma} \sqrt{\lambda_j} \leq \hat{\beta}_j - \beta_j \leq t_\delta \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left(-\hat{\beta}_j - t_\delta \hat{\sigma} \sqrt{\lambda_j} \leq -\beta_j \leq -\hat{\beta}_j + t_\delta \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left(\hat{\beta}_j + t_\delta \hat{\sigma} \sqrt{\lambda_j} \geq \beta_j \geq \hat{\beta}_j - t_\delta \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left(\hat{\beta}_j - t_\delta \hat{\sigma} \sqrt{\lambda_j} \leq \beta_j \leq \hat{\beta}_j + t_\delta \hat{\sigma} \sqrt{\lambda_j}\right) \\ &= \mathbb{P}\left([\hat{\beta}_j - \hat{\sigma} \sqrt{\lambda_j} t_\delta, \hat{\beta}_j + \hat{\sigma} \sqrt{\lambda_j} t_\delta]\right) \\ &= \mathbb{P}(\kappa_j \ni \beta_j) \end{aligned} \quad (7.20)$$

und damit ist alles gezeigt. □

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Wie gewohnt betrachten wir als erstes Beispiel die ALM Form des Szenarios unabhängig und identisch normalverteilter Zufallsvariablen

$$v \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^n, \beta := \mu \in \mathbb{R}, \sigma^2 > 0 \quad (7.21)$$

Dann gelten, wie bereits gesehen

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n v_i =: \bar{v}, \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2 =: s_v^2 \text{ und } \lambda_1 = (\mathbf{1}_n^T \mathbf{1}_n)^{-1} = \frac{1}{n}. \quad (7.22)$$

Nach dem Theorem zu Konfidenzintervallen für Betaparameterkomponenten gilt dann, dass

$$\kappa := \left[\bar{v} - \frac{s}{\sqrt{n}} t_\delta, \bar{v} + \frac{s}{\sqrt{n}} t_\delta \right] \quad (7.23)$$

ein δ -Konfidenzintervall für β ist und dieses ist offenbar identisch mit dem bekannten δ -Konfidenzintervall für den Erwartungsparameter der Normalverteilung.

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 δ -Konfidenzintervalls erinnern: Realisierungen von δ -Konfidenzintervallen überdecken den wahren, aber unbekanntem, Parameterwert mit einer frequentistischen Wahrscheinlichkeit von δ . Abbildung 7.3 zeigt, 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.

```

1 # Modellformulierung
2 library(MASS) # multivariate Normalverteilung
3 set.seed(0) # random number generator seed
4 ns = 1e2 # Anzahl Simulationen
5 n = 10 # Anzahl von Datenpunkten
6 p = 2 # Anzahl von Betaparametern
7 x = 1:n # Prädiktorwerte
8 X = matrix(c(rep(1,n),x), ncol = p) # Designmatrix
9 I_n = diag(n) # Einheitsmatrix
10 beta = matrix(c(1,2), nrow = 2) # wahre, aber unbekannte, Betaparameter
11 sigsq = 1 # wahrer, aber unbekannter, Varianzparameter
12 delta = 0.95 # Konfidenzbedingung
13 t_delta = qt((1+delta)/2,n-1) # \Psi^{-1}((1+\delta)/2,n-1)
14 lambda = diag(solve(t(X) %*% X)) # \lambda_j Werte
15
16 # Simulation
17 kappa = array(rep(NA, ns*p*p), dim=c(ns,2,2)) # Konfidenzintervallarray
18 beta_hat = matrix(rep(NA,p*ns), nrow = p) # Betaparameterschätzer
19 for(i in 1:ns){ # Iteration über Realisierungen
20   y = mvrnorm(1, X %*% beta, sigsq*I_n) # Datenrealisierung
21   beta_hat[,i] = solve(t(X) %*% X) %*% t(X) %*% y # \hat{\beta}
22   eps_hat = y - X %*% beta_hat[,i] # \hat{\varepsilon}
23   sigsq_hat = (t(eps_hat) %*% eps_hat)/(n-p) # \hat{\sigma}^2
24   for(j in 1:p){ # Iteration über Betaarraykomponenten
25     kappa[i,1,j] = beta_hat[j,i]-sqrt(sigsq_hat*lambda[j])*t_delta # untere KI Grenze
26     kappa[i,2,j] = beta_hat[j,i]+sqrt(sigsq_hat*lambda[j])*t_delta # obere KI Grenze
27   }
28 }

```

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 (1925b), Fisher (1925c) 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 der T-Statistik wieder.
2. Erläutern Sie für die T-Statistik die Bedeutung der Wahl von $c \in \mathbb{R}^p$.
3. Erläutern Sie für die T-Statistik die Bedeutung der Wahl von $\beta_0 \in \mathbb{R}^p$.
4. Wann kann die T-Statistik als Signal-zu-Rauschen Verhältnis interpretiert werden?
5. Geben Sie das Theorem zur T-Statistik wieder.
6. Geben Sie die Form der T-Statistik im Szenario von n unabhängig und identisch normalverteilten Zufallsvariablen wieder.
7. Geben Sie das Theorem zu Konfidenzintervallen für Betaparameterkomponenten wieder.

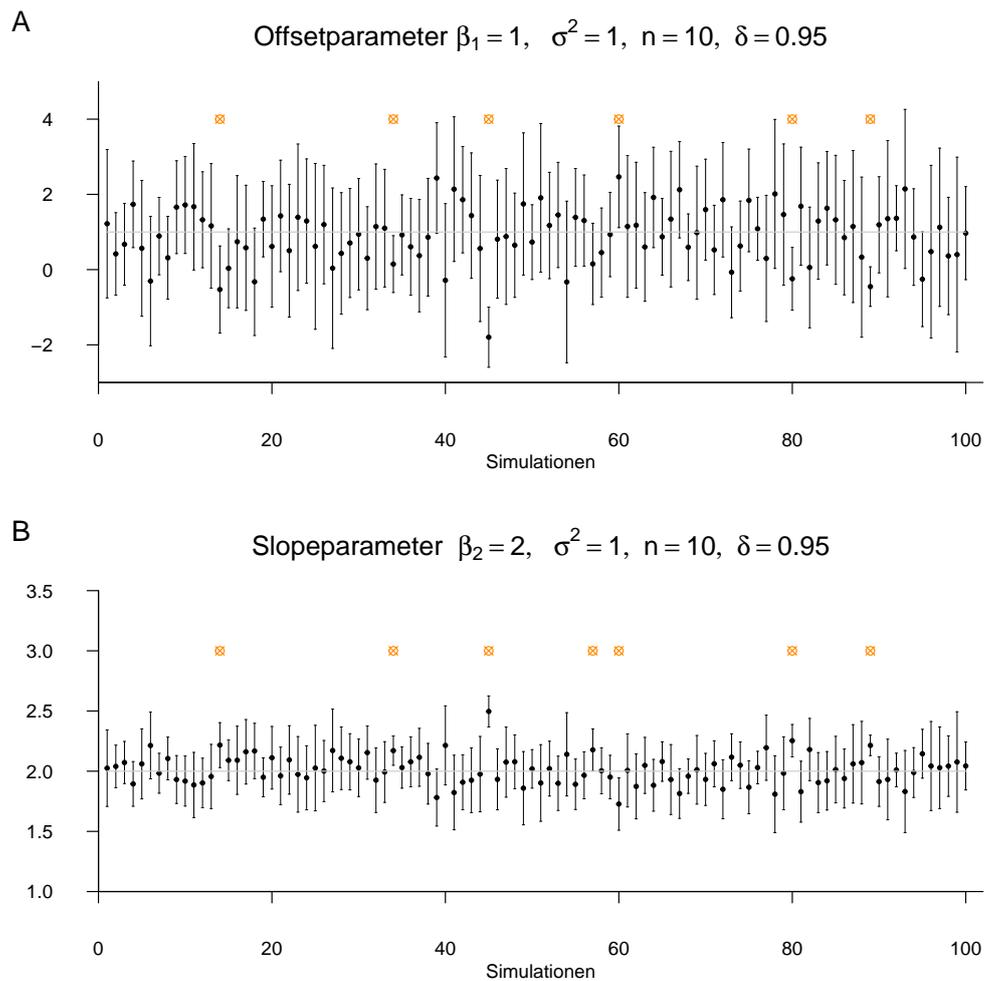


Abbildung 7.3 Konfidenzintervalle für Betaparameterkomponenten bei einfacher linearer Regression.

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 dabei 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

$$\mathcal{M}_0 := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_{\theta_0}^0 | \theta_0 \in \Theta_0\}) \text{ und } \mathcal{M}_1 := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_{\theta_1}^1 | \theta_1 \in \Theta_1\}) \quad (8.1)$$

mit identischem Datenraum, identischer σ -Algebra und potentiell distinkten Wahrscheinlichkeitsmaßmengen und Parameterräumen. Sei weiterhin v ein Zufallsvektor mit Datenraum \mathcal{Y} . Seien schließlich L_0^v und L_1^v die Likelihood-Funktionen von \mathcal{M}_0 und \mathcal{M}_1 , respektive, wobei das Superskript v jeweils an die Datenabhängigkeit der Likelihood Funktion erinnern soll. Dann wird

$$\Lambda := \frac{\max_{\theta_0 \in \Theta_0} L_0^v(\theta_0)}{\max_{\theta_1 \in \Theta_1} L_1^v(\theta_1)}, \quad (8.2)$$

Likelihood-Quotienten-Statistik genannt.

•

Eine Likelihood-Quotienten-Statistik setzt die Wahrscheinlichkeitsmasse bzw. Wahrscheinlichkeitsdichte eines beobachteten Datensatzes $y \in \mathcal{Y}$ unter zwei statistischen Modellen nach Optimierung der jeweiligen Modellparameter ins Verhältnis. Ein hoher Wert der Likelihood-Quotienten-Statistik entspricht einer höheren Wahrscheinlichkeitsmasse bzw. Wahrscheinlichkeitsdichte des beobachteten Datensatzes $y \in \mathcal{Y}$ unter \mathcal{M}_0 als unter \mathcal{M}_1 und vice versa.

Die Wahrscheinlichkeitsmassen bzw. 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 Genese 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 Factors oder der des Bayesian Information Criteria 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 , respektive, 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 nennen wir

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (8.5)$$

das *vollständige Modell* und

$$v = X_0\beta_0 + \varepsilon_0 \text{ mit } \varepsilon_0 \sim N(0_n, \sigma^2 I_n) \quad (8.6)$$

das *reduzierte Modell* und sprechen von einer *Partitionierung eines (vollständigen) Modells*. •

Man sagt auch, dass das reduzierte Modell im vollständigen Modell *geschachtelt (nested)* ist. Die Likelihood-Quotienten-Statistik beim Vergleich eines vollständigen und eines reduzierten Modells hat eine einfache Form. Diese ist der zentrale Aspekt folgenden Theorems.

Theorem 8.1 (Likelihood-Quotienten-Statistik von vollständigem und reduziertem Modell). Für $p = p_0 + p_1, p > 1$ 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 und reduziertem Modell, respektive. 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

$$\hat{\sigma}^2 = \frac{1}{n} (v - X\hat{\beta})^T (v - X\hat{\beta}) \quad \text{und} \quad \hat{\sigma}_0^2 = \frac{1}{n} (v - X_0\hat{\beta}_0)^T (v - X_0\hat{\beta}_0) \quad (8.8)$$

respektive, gegeben sind, wobei $\hat{\beta}$ und $\hat{\beta}_0$ die Maximum-Likelihood Schätzer der Betaparameter unter vollständigem und reduziertem Modell, respektive, bezeichnen. Weiterhin halten wir fest, dass für die Likelihood-Funktion des vollständigen Modells an der Stelle der Maximum-Likelihood Schätzer gilt, dass

$$\begin{aligned} L_1^y(\hat{\beta}, \hat{\sigma}^2) &= (2\pi)^{-\frac{n}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}^2} (y - X\hat{\beta})^T (y - X\hat{\beta})\right) \\ &= (2\pi)^{-\frac{n}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2} \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{(y - X\hat{\beta})^T (y - X\hat{\beta})}\right) \\ &= (2\pi)^{-\frac{n}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned} \quad (8.9)$$

und analog, dass für die Likelihood-Funktion des reduzierten Modells an der Stelle der Maximum-Likelihood Schätzer gilt, dass

$$L_0^y(\hat{\beta}_0, \hat{\sigma}_0^2) = (2\pi)^{-\frac{n}{2}} (\hat{\sigma}_0^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \quad (8.10)$$

Damit ergibt sich dann aber

$$\Lambda = \frac{\max_{\theta_0 \in \Theta_0} L_0^v(\theta_0)}{\max_{\theta_1 \in \Theta_1} L_1^v(\theta_1)} = \frac{L_0^v(\hat{\beta}_0, \hat{\sigma}_0^2)}{L_1^v(\hat{\beta}, \hat{\sigma}^2)} = \frac{(2\pi)^{-\frac{n}{2}} (\hat{\sigma}_0^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}}{(2\pi)^{-\frac{n}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \right)^{-\frac{n}{2}} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{\frac{n}{2}}$$

□

8.2 Definition und Verteilung

Wir definieren nun die F-Statistik vor dem Hintergrund eines vollständigen und eines reduzierten Modells.

Definition 8.3. Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$v = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (8.11)$$

mit der Partitionierung

$$X = (X_0 \quad X_1), \quad X_0 \in \mathbb{R}^{n \times p_0}, \quad X_1 \in \mathbb{R}^{n \times p_1}, \quad \text{und} \quad \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \beta_0 \in \mathbb{R}^{p_0}, \quad \beta_1 \in \mathbb{R}^{p_1}, \quad (8.12)$$

mit $p = p_0 + p_1$ gegeben. Weiterhin seien mit

$$\hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T v \quad \text{und} \quad \hat{\beta} := (X^T X)^{-1} X^T v \quad (8.13)$$

die Residuenvektoren

$$\hat{\varepsilon}_0 := v - X_0 \hat{\beta}_0 \quad \text{und} \quad \hat{\varepsilon} := v - X \hat{\beta} \quad (8.14)$$

definiert. Dann ist die F-Statistik definiert als

$$F := \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \quad (8.15)$$

•

Der Zähler der F-Statistik

$$\frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_1} \quad (8.16)$$

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 Größe der Residualquadratsummenreduktion (und gleichem Nenner) ein größerer F Wert resultiert, wenn diese durch weniger zusätzliche Regressoren resultiert, also p_1 klein ist (und vice versa). Im Sinne der Anzahl der Spalten von X und der entsprechenden Komponenten von β favorisiert die F -Statistik also weniger “komplexe” Modelle.

Für den Nenner der F-Statistik gilt

$$\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - p} = \hat{\sigma}^2, \quad (8.17)$$

wobei $\hat{\sigma}^2$ hier der aufgrund des vollständigen Modells geschätzte Schätzer von σ^2 ist. Werden die Daten tatsächlich unter dem reduzierten Modell generiert, so kann das vollständige Modell dies durch $\hat{\beta}_2 \approx 0_{p_1}$ abbilden und erreicht eine ähnliche σ^2 Schätzung wie das reduzierte Modell. Werden die Daten de-facto unter dem vollständigem Modell generiert, so ist $\hat{\varepsilon}^T \hat{\varepsilon} / (n - p)$ ein besserer Schätzer von σ^2 als $\hat{\varepsilon}_0^T \hat{\varepsilon}_0 / (n - p)$, da sich für diesen die Datenvariabilität, die nicht durch die p_0 Regressoren in X_0 erklärt wird, in der Schätzung von σ^2 widerspiegeln würde. Der Nenner der F-Statistik ist also in beiden Fällen der sinnvollere Schätzer von σ^2 .

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 (σ^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

$$X = (X_0 \ X_1), X_0 := 1_n, X_1 := (x_1, \dots, x_n)^T, \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad (8.18)$$

und zwar einmal für den Fall, dass $\beta = (1, 0)^T$, also dass das reduzierte Modell das wahre, aber unbekannte, datenerzeugende Modell ist und einmal für den Fall, dass $\beta = (1, 0.5)^T$, also dass 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.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 nmod = 2 # Anzahl Modelle
4 n = 10 # Anzahl Datenpunkte
5 p = 2 # Anzahl Betaparameter
6 p_0 = 1 # Anzahl Betaparameter reduziertes Modell
7 p_1 = 1 # Anzahl zusätzlicher Betaparameter vollständiges Modell
8 p = p_0 + p_1 # Anzahl Betaparameter im vollständigem Modell
9 x = 1:n # Prädiktorwerte
10 X = matrix(c(rep(1,n),x), nrow = n) # Designmatrix des vollständigen Modells
11 X_0 = X[,1] # Designmatrix des reduzierten Modells
12 I_n = diag(n) # n x n Einheitsmatrix
13 beta = matrix(c(1,0,1,.5), nrow = 2) # wahre , aber unbekannte , Betaparameter
14 nscn = ncol(beta) # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
15 sigsqr = 1 # wahrer, aber unbekannter, Varianzparameter
16
17 # Modellsimulation und Evaluierung
18 Eff = matrix(rep(NA, nscn), nrow = nscn) # F-Statistik Realisierungsarray
19 for(s in 1:nscn){ # Szenarieniterationen
20 y = mvrnorm(1, X %*%beta[s], sigsqr*I_n) # Datenrealisierung

```

```

21 beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y # Betaparameterschätzer reduziertes Modell
22 beta_hat = solve(t(X) %*%X )%*%t(X) %*%y # Betaparameterschätzer vollständiges Modell
23 eps_0_hat = y-X_0%*%beta_hat_0 # Residuenvektor reduziertes Modell
24 eps_hat = y-X%*%beta_hat # Residuenvektor vollständiges Modell
25 eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat # RQS reduziertes Modell
26 eps_eps_hat = t(eps_hat) %*% eps_hat # RQS vollständiges Modell
27 Eff [s] = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/ # F-Statistik
28 (eps_eps_hat/(n-p))))

```

F-Statistik für beta_1 = 0_{p_1}: 0.6616985
 F-Statistik für beta_1 != 0_{p_1}: 10.50031

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 Λ seien die entsprechenden F- und Likelihood-Quotienten-Statistiken. Dann gilt*

$$F = \frac{n-p}{p_1} \left(\Lambda^{-\frac{2}{n}} - 1 \right). \quad (8.19)$$

Beweis. Wir erinnern zunächst daran, dass die Maximum-Likelihood Schätzer des Varianzparameters durch

$$\hat{\sigma}^2 = \frac{1}{n} (v - X\hat{\beta})^T (v - X\hat{\beta}) = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n} \quad \text{und} \quad \hat{\sigma}_0^2 = \frac{1}{n} (v - X_0\hat{\beta}_0)^T (v - X_0\hat{\beta}_0) = \frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0}{n} \quad (8.20)$$

gegeben sind. Mit der Definition der F-Statistik und der Form der Likelihood-Quotienten-Statistik für den Vergleich von reduziertem und vollständigem Modell ergibt sich dann

$$\begin{aligned}
 F &= \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \\
 &= \frac{n(\hat{\sigma}_0^2 - \hat{\sigma}^2)/p_1}{n\hat{\sigma}^2/(n-p)} \\
 &= \frac{n-p}{p_1} \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \\
 &= \frac{n-p}{p_1} \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1 \right) \\
 &= \frac{n-p}{p_1} \left(\Lambda^{-\frac{2}{n}} - 1 \right)
 \end{aligned} \quad (8.21)$$

□

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.

Wir dokumentieren die Frequentistische Verteilung der F-Statistik in folgendem Theorem, auf dessen Beweis wir verzichten.

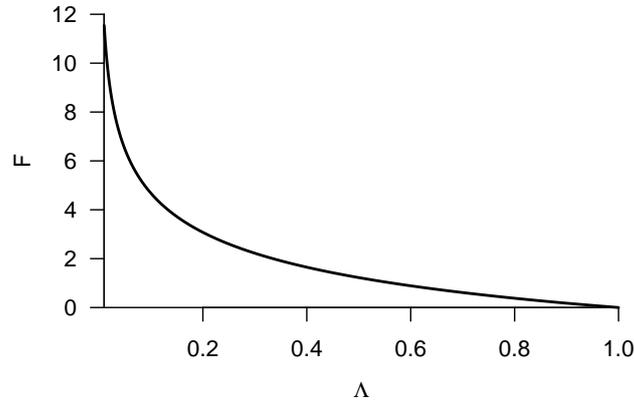


Abbildung 8.1 Zusammenhang von F- und Likelihood-Quotienten-Statistik.

Theorem 8.3 (F-Statistik). Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (8.22)$$

mit der Partitionierung

$$X = (X_0 \ X_1), X_0 \in \mathbb{R}^{n \times p_0}, X_1 \in \mathbb{R}^{n \times p_1}, \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \beta_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}, \quad (8.23)$$

mit $p = p_0 + p_1$ gegeben. Schließlich sei

$$c := \begin{pmatrix} 0_{p_0} \\ 1_{p_1} \end{pmatrix} \in \mathbb{R}^p \quad (8.24)$$

ein Vektor. Dann gilt

$$F \sim f(\delta, p_1, n - p) \text{ mit } \delta := \frac{c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta}{\sigma^2} \quad (8.25)$$

Man beachte, dass die F -Statistik eine Funktion des Parameterschätzers, δ dagegen eine Funktion der wahren, aber unbekannt, Parameter ist. Wie die Verteilung der T -Statistik kann 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 von Gleichung 8.18 und zwar erneut einmal für den Fall, dass $\beta = (1, 0)^T$, also dass das reduzierte Modell das wahre, aber unbekannt, datenerzeugende Modell ist und einmal für den Fall, dass $\beta = (1, 0.5)^T$, also dass das vollständige Modell das wahre, aber unbekannt, datenerzeugende Modell ist. Abbildung 8.2 A und B visualisieren die resultierenden Verteilungen, respektive.

```

1 # Modellformulierung
2 library(MASS)
3 nmod = 2
4 n = 10
5 p_0 = 1
6 p_1 = 1
7 p = p_0 + p_1
8 x = 1:n
9 X = matrix(c(rep(1,n),x), nrow = n)
10 X_0 = X[,1]
11 I_n = diag(n)
12 beta = matrix(c(1,0,1,.5), nrow = 2)
13 nscn = ncol(beta)
14 sigsq = 1
15 c = matrix(c(0,1), nrow = 2)
16
17 # Frequentistische Simulation
18 nsim = 1e4
19 delta = rep(NA,nscn)
20 Eff = matrix(rep(NA,n), nscn*nsim), nrow = nscn)
21 for(s in 1:nscn){
22   delta[s] = (t(c)%*%beta[,s])%*%
23             (solve(t(c)%*%solve(t(X)%*%X)%*%c) %*%
24              (t(c)%*%beta[,s])/sigsq)
25   for(i in 1:nsim){
26     y = mvrnorm(1, X %*%beta[,s], sigsq*I_n)
27     beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y
28     beta_hat = solve(t(X) %*%X) %*%t(X) %*%y
29     eps_0_hat = y-X_0%*%beta_hat_0
30     eps_hat = y-X%*%beta_hat
31     eps_0_eps_0_hat = t(eps_0_hat) %*% eps_0_hat
32     eps_eps_hat = t(eps_hat) %*% eps_hat
33     Eff[s,i] = (((eps_0_eps_0_hat-eps_eps_hat)/p_1)/
34                (eps_eps_hat/(n-p))))}

```

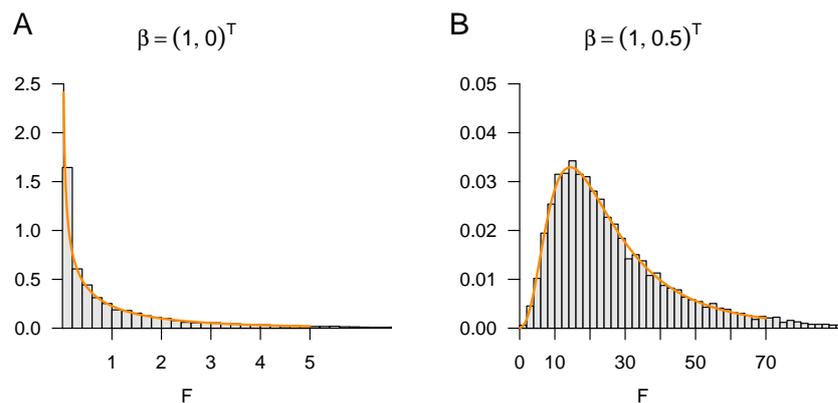


Abbildung 8.2 Exemplarische F-Statistik Verteilungen unter reduziertem und vollständigem wahren, aber unbekanntem, datenerzeugendem Modell bei einfacher linearer Regression

8.3 Literaturhinweise

Die Popularität von F-Statistiken, insbesondere im Kontext der Varianzanalyse, wird allgemein auf Fisher (1925b) 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 der Likelihood-Quotienten-Statistik wieder.

2. Erläutern Sie die Definition der Likelihood-Quotienten-Statistik.
3. Geben Sie die Definition eines vollständigem und eines reduziertem ALMs wieder.
4. Geben Sie das Theorem zum Likelihood-Quotienten von vollständigem und reduzierten ALM wieder.
5. Definieren Sie die F-Statistik.
6. Erläutern Sie den Zähler der F-Statistik.
7. Erläutern Sie den Nenner der F-Statistik.
8. Erläutern Sie die F-Statistik.
9. Geben Sie das Theorem zum Zusammenhang von F-Statistik und Likelihood-Quotienten-Statistik wieder.
10. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

9 T-Tests

9.1 Einstichproben-T-Tests

Anwendungsszenario

Das Anwendungsszenario eines Einstichproben-T-Test 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 μ als auch der Varianzparameter σ^2 unbekannt sind. Schließlich wird vorausgesetzt, dass ein Interesse an einem inferentiellen Vergleich des unbekanntes Erwartungswertparameters μ mit einem vorgegebenen Wert μ_0 (z.B. $\mu_0 := 0$) besteht.

Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel betrachten wir die Analyse von Pre-Post-Interventions-BDI-Differenzwerten einer 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 (PosBDI) der psychotherapeutischen Intervention, die dritte Spalte dBDI zeigt die entsprechenden PreBDI-PosBDI Differenzwerte. Ein positiver Wert entspricht hier einer Verbesserung der Depressionssymptomatik und ein negativer Wert einer Verschlechterung der Depressionssymptomatik

Tabelle 9.1 Pre und Post Intervention BDI Werte

PreBDI	PosBDI	dBDI
32	18	14
37	27	10
29	30	-1
33	31	2
33	30	3
32	31	1
26	31	-5
31	27	4
37	27	10
26	32	-6
31	17	14
27	24	3

Bei der Anwendung eines Einstichproben-T-Tests auf die dBDI Daten dieses Datensatzes nehmen wir also an, dass die dBDI Daten Realisierungen von $n = 12$ unabhängig normalverteilten Zufallsvariablen $v_i \sim N(\mu, \sigma^2)$ sind. Wir nehmen weiterhin an, dass wir daran interessiert sind, unsere Unsicherheit beim inferentiellen Vergleich des wahren, aber

unbekannten, Erwartungswertparameters μ mit einem Vergleichswert μ_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 die PreBDI und PosBDI also zwar in positiver Richtung, was eine Verringerung der Depressionssymptomatik impliziert, allerdings streuen die Daten auch über Patient:innen 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
dBDI	12	14	-6	3	4.08	44.81	6.69

Modellformulierung

Wir definieren nun das Einstichproben-T-Test-Modell wie folgt.

Definition 9.1 (Einstichproben-T-Test-Modell). Für $i = 1, \dots, n$ seien v_i Zufallsvariablen, die die n Datenpunkte eines Einstichproben-T-Test-Szenarios modellieren. Dann hat das *Einstichproben-T-Test-Modell* die strukturelle Form

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.1)$$

die Datenverteilungsform

$$v_i \sim N(\mu, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.2)$$

und für den Datenvektor $v = (v_1, \dots, v_n)^T$ die Designmatrixform

$$v = X\beta + \varepsilon \text{ mit } X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ und } \sigma^2 > 0. \quad (9.3)$$

•

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 Modellformulierung 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. Unterer **R** Code demonstriert dies.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n = 40 # Anzahl von Datenpunkten
4 p = 1 # Anzahl von Betaparameter
5 X = matrix(rep(1,n), nrow = n) # Designmatrix
6 I_n = diag(n) # n x n Einheitsmatrix
7 beta = 5 # wahrer, aber unbekannter, Betaparameter
8 sigsq = 14 # wahrer, aber unbekannter, Varianzparameter
9
10 # Datenrealisierung
11 y = mvrnorm(1, X %>% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs

```

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*

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n v_i =: \bar{v}, \quad (9.4)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2 =: s_v^2 \quad (9.5)$$

\bar{v} und s_v^2 bezeichnen hier also wiederum das Stichprobenmittel und die Stichprobenvarianz der Zufallsvariablen v_1, \dots, v_n .

Modellevaluation

Basierend auf Theorem 7.1 in Kapitel 7 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

$$T = \sqrt{n} \left(\frac{\bar{v} - \mu_0}{s_v} \right) \quad (9.7)$$

und es gilt, dass

$$T \sim t(\delta, n-1) \text{ mit } \delta = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right). \quad (9.8)$$

Beweis. Mit dem T-Teststatistik Theorem in Kapitel 7 gilt

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{v} - 1^T \mu_0}{\sqrt{s_v^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left(\frac{\bar{v} - \mu_0}{s_v} \right). \quad (9.9)$$

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 Einstichproben-T-Test-Spezialfall des ALMs 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 μ sowie Durchführung eines zweiseitigen Einstichproben-T-Tests mit einfacher Nullhypothese $\Theta_0 := \{0\}$ und Signifikanzlevel $\alpha_0 := 0.05$ für das oben skizzierte Anwendungsbeispiel. Die Nullhypothese würde in diesem Fall mit einem Freiheitsgradparameter von 11, einem kritischen Wert von $k_{0.05} = 2.20$ und einem Wert der T-Statistik von $T = 2.11$ nicht verworfen werden. Das 95%-Konfidenzintervall für den wahren, aber unbekanntem, Erwartungswertparameter ist $[-0.16, 8.33]$, überdeckt also insbesondere auch den Wert $\mu = 0$.

```

1 # Dateneinlesen
2 fname = file.path("../data/409-ETT.csv") # Dateiname
3 D = read.table(fname, sep = ",", header = TRUE) # Dataframe
4 y = D$BDI
5
6 # Modellformulierung
7 n = length(y) # Anzahl Datenpunkte
8 p = 1 # Anzahl Betaparameter
9 X = matrix(rep(1,n), nrow = n) # Designmatrix
10
11 # Parameterschätzung
12 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
13 eps_hat = y - X %*% beta_hat # Residuenvektor
14 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
15
16 # Konfidenzintervall
17 delta = 0.95 # Konfidenzbedingung
18 t_delta = qt((1+delta)/2, n-1) # \Psi^{-1}((1+\delta)/2, n-1)
19 lambda = diag(solve(t(X) %*% X)) # \lambda_j Werte
20 kappa = matrix(rep(NA, p*2), nrow = p) # \beta_j Konfidenzintervall array
21 for(j in 1:p){ # Iteration über \beta_j
22   kappa[j,1] = beta_hat[j] - sqrt(sigsq_hat*lambda[j])*t_delta # untere KI Grenze
23   kappa[j,2] = beta_hat[j] + sqrt(sigsq_hat*lambda[j])*t_delta # obere KI Grenze
24 }
25 # Hypothesentest
26 c = matrix(c(1), nrow = p) # Kontrastgewichtsvektor
27 mu_0 = 0 # Nullhypothese H_0
28 alpha_0 = 0.05 # Signifikanzniveau
29 k_alpha_0 = qt(1 - (alpha_0/2), n-1) # kritischer Wert
30 t_num = t(c) %*% beta_hat - mu_0 # T-Teststatistik Zähler
31 t_den = sqrt(sigsq_hat %*% t(c)*solve(t(X) %*% X)%*%c) # T-Teststatistik Nenner
32 t = t_num/t_den # T-Teststatistik
33 if(abs(t) >= k_alpha_0){phi = 1} else {phi = 0} # Test 1_{|T(X)| >= k_alpha_0}
34 pval = 2*(1 - pt(abs(t), n-1)) # p-Wert
35
36 # Ausgabe
37 cat("fg = ", n-1,
38     "\nkappa = ", kappa,
39     "\nt = ", t,
40     "\nalpha_0 = ", alpha_0,
41     "\nk_alpha_0 = ", k_alpha_0,
42     "\nphi = ", phi,
43     "\np-Wert = ", pval)

```

```

fg = 11
kappa = -0.1698774 8.336544
T = 2.113076
alpha_0 = 0.05
k_alpha_0 = 2.200985
phi = 0
p-Wert = 0.05826542

```

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 μ_1 und Varianzparameter σ^2 sind, während weiterhin angenommen wird, dass die n_2 univariaten Datenpunkte der zweiten Gruppe Realisierungen von n_2 unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter μ_2 und Varianzparameter σ^2 sind. Es wird also insbesondere angenommen, dass sich die wahren, aber unbekannt, Erwartungswertparameter beider Gruppen von Zufallsvariablen unterscheiden können, die Varianzparameter beider Gruppen dagegen werden als identisch angenommen. Schließlich wird vorausgesetzt, dass ein Interesse am inferentiellen Vergleich der unbekannt Erwartungswertparameter μ_1 und μ_2 besteht, so zum Beispiel ihrer Gleichheit $\mu_1 = \mu_2$ oder Verschiedenheit $\mu_1 \neq \mu_2$.

Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel betrachten wird die Analyse von Pre-Post-Interventions-BDI-Differenzwerten zweier Gruppen von je 12 Patient:innen in unterschiedlichen Therapiesettings, wie in Tabelle Tabelle 9.3 dargestellt. Die erste Spalte der Tabelle (COND) listet das patientenspezifische Therapiesetting (F2F: face-to-face, ONL: online) auf. Die zweite Spalte der Tabelle (dBDI) listet die entsprechenden patientenspezifischen Pre-Post-Interventions-BDI-Differenzwerte auf. Positive Werte entsprechen hier erneut einer Abnahme der Depressionssymptomatik, negative Werte einer Zunahme der Depressionssymptomatik.

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 $v_{1j} \sim N(\mu_1, \sigma^2)$ 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 $v_{2j} \sim N(\mu_2, \sigma^2)$ 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 Tabelle 9.4 aufgeführt.

Tabelle 9.3 Exemplarische Pre-Post-BDI Differenzwerte für zwei Stichproben.

COND	dBDI
F2F	8
F2F	14
F2F	16
F2F	17
F2F	16
F2F	16
F2F	17
F2F	14
F2F	14
F2F	17

COND	dBDI
F2F	8
F2F	12
ONL	11
ONL	14
ONL	9
ONL	12
ONL	12
ONL	11
ONL	8
ONL	11
ONL	15
ONL	8
ONL	11
ONL	8

Tabelle 9.4 Deskriptivstatistiken der Pre-Post-Interventions-BDI-Differenzwerte bei unterschiedlichen Therapiesettings

	n	Max	Min	Median	Mean	Var	Std
F2F	12	17	8	15	14.08	10.45	3.23
ONL	12	15	8	11	10.83	5.24	2.29

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). Für $i = 1, 2$ und $j = 1, \dots, n_i$ seien v_{ij} Zufallsvariablen, die die $n = n_1 + n_2$ Datenpunkte eines Zweistichproben-T-Test Szenarios modellieren. Dann hat das *Zweistichproben-T-Test-Modell* die strukturelle Form

$$v_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.11)$$

die Datenverteilungsform

$$v_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \quad (9.12)$$

und für den n -dimensionalen Datenvektor definiert als

$$v := (v_{11}, \dots, v_{1n_1}, v_{21}, \dots, v_{2n_2})^T \quad (9.13)$$

die Designmatrixform

$$v = X\beta + \varepsilon \quad (9.14)$$

mit

$$X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (9.15)$$

•

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-Modell 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.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 n_1 = 40 # Anzahl von Datenpunkten Gruppe 1
4 n_2 = 40 # Anzahl von Datenpunkten Gruppe 2
5 n = n_1 + n_2 # Gesamtanzahl Datenpunkte
6 p = 2 # Anzahl von Betaparameter
7 X = matrix(c(rep(1,n_1), rep(0,n_1), # Designmatrix
8             rep(0,n_2), rep(1,n_2)),
9            nrow = n)
10 I_n = diag(n) # n x n Einheitsmatrix
11 beta = matrix(c(1,2), nrow = p) # wahrer, aber unbekannter, Betaparameter
12 sigsq = 14 # wahrer, aber unbekannter, Varianzparameter
13
14 # Datenrealisierung
15 y = mvrnorm(1, X %>% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs

```

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 *gepoolte 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*

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} v_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix} \quad (9.16)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (v_{1j} - \bar{v}_1)^2 + \sum_{j=1}^{n_2} (v_{2j} - \bar{v}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \quad (9.17)$$

Beweis. Für $i = 1, 2$ sei $v_i := (v_{i1}, \dots, v_{in_i})^T$. Dann ergibt sich für den Betaparameterschätzer

$$\begin{aligned}
 \hat{\beta} &= (X^T X)^{-1} X^T y \\
 &= \left(\begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\
 &= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} v_{1j} \\ \sum_{j=1}^{n_2} v_{2j} \end{pmatrix} \\
 &= \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} v_{1j} \\ \sum_{j=1}^{n_2} v_{2j} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} v_{2j} \end{pmatrix} \\
 &=: \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix}.
 \end{aligned} \tag{9.18}$$

Gleichsam ergibt sich für Varianzparameterschätzer mit $n = n_1 + n_2$ und $p = 2$

$$\begin{aligned}
 \hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\
 &= \frac{1}{n_1 + n_2 - 2} \left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix} \right)^T \left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix} \right) \\
 &= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} v_{11} - \bar{v}_1 \\ \vdots \\ v_{1n_1} - \bar{v}_1 \\ v_{21} - \bar{v}_2 \\ \vdots \\ v_{2n_2} - \bar{v}_2 \end{pmatrix}^T \begin{pmatrix} v_{11} - \bar{v}_1 \\ \vdots \\ v_{1n_1} - \bar{v}_1 \\ v_{21} - \bar{v}_2 \\ \vdots \\ v_{2n_2} - \bar{v}_2 \end{pmatrix} \\
 &= \frac{\sum_{j=1}^{n_1} (v_{1j} - \bar{v}_1)^2 + \sum_{j=1}^{n_2} (v_{2j} - \bar{v}_2)^2}{n_1 + n_2 - 2} \\
 &=: s_{12}^2.
 \end{aligned} \tag{9.19}$$

□

Man beachte, dass sich die Stichprobenvarianz s_v^2 der Komponenten von v im Allgemeinen von der gepoolten Stichprobenvarianz s_{12}^2 unterscheidet. Dies ist nicht zuletzt dadurch bedingt, dass die Stichprobenvarianz basierend auf dem Gesamtstichprobenmittel \bar{v} , die gepoolte Stichprobenvarianz dagegen basierend auf den gruppenspezifischen Stichprobenmittel \bar{v}_1 und \bar{v}_2 ermittelt wird. Wir wollen das Konzept der gepoolten Stichprobenvarianz hier aber nicht weiter vertiefen.

Modellevaluation

Basierend auf Theorem 7.1 in Kapitel 7 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-Tests. Dann ergibt sich für die T-Teststatistik mit*

$$c := (1, -1)^T \text{ und } c^T \beta_0 =: \mu_0, \tag{9.20}$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{v}_1 - \bar{v}_2 - \mu_0}{s_{12}} \right) \quad (9.21)$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (9.22)$$

Beweis. Mit dem T-Statistik Theorem Kapitel 7 Modellevaluation gilt zunächst für die Zähler von T und δ , dass

$$c^T \hat{\beta} - c^T \beta_0 = (1 \quad -1) \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix} - \mu_0 = \bar{v}_1 - \bar{v}_2 - \mu_0 \quad (9.23)$$

und

$$c^T \beta - c^T \beta_0 = (1 \quad -1) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0, \quad (9.24)$$

respektive. Weiterhin gilt für die Nenner von T und δ , dass

$$c^T (X^T X)^{-1} c = (1 \quad -1) \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = (n_1^{-1} \quad -n_2^{-1}) \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2} \quad (9.25)$$

Außerdem gilt

$$\left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_1 + n_2}{n_1 n_2} \right)^{-\frac{1}{2}} = \left(\frac{n_1 n_2}{n_1 + n_2} \right)^{\frac{1}{2}} \quad (9.26)$$

Zusammengenommen folgt direkt, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{v}_1 - \bar{v}_2 - \mu_0}{s_{12}} \right) \text{ und } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (9.27)$$

□

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 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 μ_1 und μ_2 sowie Durchführung eines zweiseitigen Zweistichproben-T-Tests mit Nullhypothese

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2 \mid \mu_1 = \mu_2 \right\} \quad (9.28)$$

und Signifikanzlevel $\alpha_0 := 0.05$ für das oben skizzierte Anwendungsbeispiel. Die Nullhypothese würde in diesem Fall mit einem Freiheitsgradparameter von 22, einem kritischen Wert von $k_{0.05} = 2.07$ und einem Wert der T-Statistik von $T = 2.84$ verworfen werden. Inferenzstatistisch besteht also Evidenz dafür, dass sich der wahre, aber unbekannte Erwartungswertparameter im F2F Therapiesetting in positiver Richtung vom wahren, aber unbekanntem Erwartungswertparameter im ONL Therapiesetting unterscheidet. Die 95%-Konfidenzintervalle für die wahren, aber unbekanntem, Erwartungswertparameter μ_1 und μ_2 sind [12.4, 15.8] und [9.2, 12.5], respektive.

```

1 # Dateneinlesen
2 fname = file.path("../_data/409-ZTT.csv") # Dateiname
3 D = read.table(fname, sep = ",", header = TRUE) # Dataframe
4 y_1 = D$BDI[D$COND == "F2F"] # BDI Differenzwerte in der F2F Gruppe
5 y_2 = D$BDI[D$COND == "ONL"] # BDI Differenzwerte in der ONL Gruppe
6
7 # Modellformulierung
8 n_1 = length(y_1) # Anzahl Datenpunkte Gruppe 1 (F2F)
9 n_2 = length(y_2) # Anzahl Datenpunkte Gruppe 2 (ONL)
10 n = n_1 + n_2 # Gesamtanzahl Datenpunkte
11 y = matrix(c(y_1, y_2), nrow = n) # Datenvektor
12 p = 2 # Anzahl Betaparameter
13 X = matrix(c(rep(1,n_1), rep(0,n_2), # Zweistichproben-T-Test Designmatrix
14 rep(0,n_1), rep(1,n_2)),
15 nrow = n)
16 # Parameterschätzung
17 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
18 eps_hat = y - X %*% beta_hat # Residuenvektor
19 sigsq_hat = t(eps_hat) %*% eps_hat / (n-p) # Varianzparameterschätzer
20
21 # Konfidenzintervall
22 delta = 0.95 # Konfidenzbedingung
23 t_delta = qt((1+delta)/2, n-1) # \Psi^{-1}((1+\delta)/2, n-1)
24 lambda = diag(solve(t(X) %*% X)) # \lambda_j Werte
25 kappa = matrix(rep(NA, p*2), nrow = p) # \beta_j Konfidenzintervall array
26 for(j in 1:p){ # Iteration über \beta_j
27 kappa[j,1] = beta_hat[j] - sqrt(sigsq_hat*lambda[j])*t_delta # untere KI Grenze
28 kappa[j,2] = beta_hat[j] + sqrt(sigsq_hat*lambda[j])*t_delta # obere KI Grenze
29
30 # Hypothesentest
31 c = matrix(c(1,-1), nrow = 2) # Kontrastgewichtsvektor
32 mu_0 = 0 # Nullhypothese H_0
33 alpha_0 = 0.05 # Signifikanzniveau
34 k_alpha_0 = qt(1 - (alpha_0/2), n-1) # kritischer Wert
35 t_num = t(c) %*% beta_hat - mu_0 # T-Teststatistik Zähler
36 t_den = sqrt(sigsq_hat*t(c) %*% solve(t(X) %*% X) %*% c) # T-Teststatistik Nenner
37 t = t_num/t_den # T-Teststatistik
38 if(abs(t) >= k_alpha_0){phi = 1} else {phi = 0} # Test 1_{|T(X)| >= k_alpha_0}
39 pval = 2*(1-pt(abs(t), n_1+n_2-2)) # p-Wert
40
41 # Ausgabe
42 cat("fg = ", n_1 + n_2 - 2,
43 "\nkappa_1 = ", kappa[1,],
44 "\nkappa_2 = ", kappa[2,],
45 "\nt = ", t,
46 "\nalpha_0 = ", alpha_0,
47 "\nk_alpha_0 = ", k_alpha_0,
48 "\nphi = ", phi,
49 "\np-Wert = ", pval)

```

```

fg = 22
kappa_1 = 12.41076 15.75591
kappa_2 = 9.160756 12.50591
T = 2.842306
alpha_0 = 0.05
k_alpha_0 = 2.068658
phi = 1
p-Wert = 0.009477564

```

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 (1925b) und Fisher (1935) verwiesen.

9.4 Selbstkontrollfragen

1. Geben Sie die Definition des Einstichproben-T-Test-Modells wieder.
2. Geben Sie das Theorem zur Parameterschätzung im Einstichproben-T-Test-Modell wieder.
3. Geben Sie das Theorem zur T-Teststatistik des Einstichproben-T-Tests wieder.
4. Geben Sie die Definition des Zweistichproben-T-Test-Modells wieder.

5. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test-Modell wieder.
6. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.

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 *iten* Gruppe bzw. des *iten* Faktorlevels wird dabei angenommen, dass sie Realisierungen von jeweils n_i unabhängigen und identisch normalverteilten Zufallsvariablen sind, deren wahre, aber unbekannte, 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 Einstich- 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, respektive, aufweist. Schließlich wird wie im Falle der T-Test Szenarien meist vorausgesetzt, dass ein Interesse an einem inferentiellen Vergleich der wahren, aber unbekanntem, faktorlevelspezifischen Erwartungswertparameter besteht.

Anwendungsbeispiel

Als konkretes Anwendungsbeispiel betrachten wir die Analyse von Pre-Post-Interventions-BDI-Differenzwerten 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:innenspezifische Therapiesetting (F2F: face-to-face, ONL:online, WLC: waitlist control) für jeweils drei Patient:innen jeder Studiengruppe auf. Die zweite Spalte der Tabelle (dBDI) list die entsprechenden patient:innenspezifischen Pre-Post-Interventions-BDI-Differenzwerte. Positive Werte entsprechen hier wieder einer Abnahme der Depressionssymptomatik, negative Werte einer Zunahme der Depressionssymptomatik.

Tabelle 10.1 Exemplarische Pre-Post-Intervention-BDI-Differenzwerte des Beispieldatensatzes

	COND	dBDI
1	F2F	9
2	F2F	6
3	F2F	10
13	ONL	1
14	ONL	7
15	ONL	9
25	WLC	-1
26	WLC	2

	COND	dBDI
27	WLC	-3

Abbildung 10.1 zeigt eine Visualisierung des gesamten Datensatzes. Die Balken repräsentieren die gruppenspezifischen Stichprobenmittelwerte, die zugehörigen Fehlerbalken die gruppenspezifischen Stichprobenstandardabweichungen. Die Punktwolken repräsentieren die gruppenspezifischen Datenpunkte. In den F2F und ONL Gruppen ist die Veränderung des BDI Wertes stärker ausgeprägt als in der WLC Gruppe.

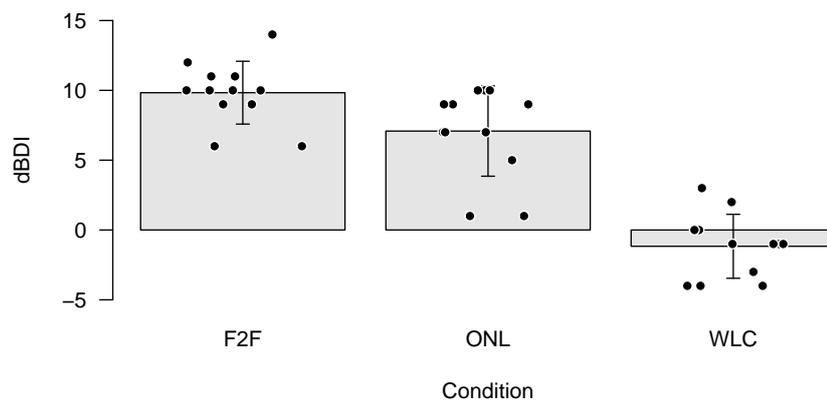


Abbildung 10.1 Datendarstellung des Anwendungsbeispiels zur Einfaktoriellen Varianzanalyse

In Tabelle 10.2 fassen wir die Deskriptivstatistiken des Beispieldatensatzes, aufgeschlüsselt nach Therapiebedingungen, zusammen.

Tabelle 10.2 Deskriptivstatistiken der Pre-Post BDI Differenzwerte.

	n	Max	Min	Median	Mean	Var	Std
F2F	12	14	6	10	9.83	5.06	2.25
ONL	12	10	1	8	7.08	10.45	3.23
WLC	12	3	-4	-1	-1.17	5.24	2.29

10.2 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 v_{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

$$v_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. mit } \mu_i \in \mathbb{R}, \sigma^2 > 0, \tag{10.1}$$

die Datenverteilungsform

$$v_{ij} \sim N(\mu_i, \sigma^2) \text{ u.v. mit } \mu_i \in \mathbb{R}, \sigma^2 > 0. \quad (10.2)$$

und für den n -dimensionalen Datenvektor definiert als

$$v := (v_{11}, \dots, v_{1n_1}, v_{21}, \dots, v_{2n_2}, \dots, v_{p1}, \dots, v_{pn_p})^T \quad (10.3)$$

die Designmatrixform

$$v = X\beta + \varepsilon \quad (10.4)$$

mit

$$X := \begin{pmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_p} & 0_{n_p} & \cdots & 1_{n_p} \end{pmatrix}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \in \mathbb{R}^p, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \quad (10.5)$$

Für $n_i := m$ für alle $i = 1, \dots, p$ heißt das Modell *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 $p = 2$ für ein beliebiges $p \in \mathbb{N}$ handelt.

Motivation der Effektdarstellung

Das EVA-Modell in Erwartungswertparameterdarstellung ist ein valides Modell, auf dessen Grundlage sowohl Parameterschätzung 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 α_i die *Differenz* zwischen dem i ten Erwartungswertparameter μ_i und dem gruppenübergreifenden Erwartungswertparameter μ_0 ,

$$\alpha_i = \mu_i - \mu_0 \text{ für } i = 1, \dots, p. \quad (10.7)$$

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 μ_0 und $\alpha_i, i = 1, \dots, p$ dargestellt. Diese Darstellung ist im Allgemeinen nicht eindeutig. Zum Beispiel können die Erwartungswertparameter $\mu_1 = 3, \mu_2 = 5, \mu_3 = 6$ sowohl durch den gruppenspezifischen Erwartungswertparameter $\mu_0 = 0$ und die gruppenunspecifischen Effektparameter $\alpha_1 = 3, \alpha_2 = 5, \alpha_3 = 6$ als auch durch den gruppenunspecifischen 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 (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 der Nebenbedingung

$$\alpha_1 := 0 \quad (10.8)$$

an. 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 α_i modellieren die Differenz zwischen dem Erwartungswertparameter der i ten Gruppe und dem Erwartungswertparameter der ersten Gruppe:

$$\alpha_i = \mu_i - \mu_0 = \mu_i - \mu_1 \text{ für } i = 1, \dots, p. \quad (10.10)$$

μ_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 dagegen 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 dem Erwartungswertparameter der Referenzgruppe und dem Erwartungswertparameter 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*

$$\begin{aligned} v_{1j} &= \mu_0 + \varepsilon_{1j} && \text{mit } \varepsilon_{1j} \sim N(0, \sigma^2) \text{ u.i.v. für } j = 1, \dots, n_1 \\ v_{ij} &= \mu_0 + \alpha_i + \varepsilon_{ij} && \text{mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 2, \dots, p, j = 1, \dots, n_i \end{aligned} \quad (10.11)$$

mit

$$\alpha_i := \mu_i - \mu_1 \text{ für } i = 2, \dots, p \quad (10.12)$$

und in der entsprechenden Datenverteilungsform

$$\begin{aligned} v_{1j} &\sim N(\mu_0, \sigma^2) && \text{u.i.v. für } j = 1, \dots, n_1 \text{ mit } \mu_1 \in \mathbb{R}, \sigma^2 > 0 \\ v_{ij} &\sim N(\mu_0 + \alpha_i, \sigma^2) && \text{u.v. für } i = 2, \dots, p, j = 1, \dots, n_i \text{ mit } \alpha_i \in \mathbb{R}, \sigma^2 > 0 \end{aligned} \quad (10.13)$$

geschrieben werden.

Beweis. Nach Definition gilt

$$\mu_i = \mu_0 + \mu_i - \mu_0. \quad (10.14)$$

Die Parametrisierungen mit μ_i und mit $\mu_0 + \mu_i - \mu_0$ sind also gleich und damit äquivalent. Dann folgt aber auch

$$\mu_i = \mu_0 + (\mu_i - \mu_0) =: \mu_0 + \alpha_i \text{ für } i = 1, \dots, p \quad (10.15)$$

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. Für $i = 1, \dots, p$ und $j = 1, \dots, n_i$ seien v_{ij} Zufallsvariablen, die die $n := \sum_{i=1}^p n_i$ Datenpunkte eines EVA-Szenarios modellieren. Dann hat das *EVA-Modell in Effektdarstellung* die strukturelle Form

$$\begin{aligned} v_{1j} &= \mu_0 + \varepsilon_{1j} && \text{mit } \varepsilon_{1j} \sim N(0, \sigma^2) \text{ u.i.v. für } j = 1, \dots, n_1 \\ v_{ij} &= \mu_0 + \alpha_i + \varepsilon_{ij} && \text{mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 2, \dots, p, j = 1, \dots, n_i \end{aligned} \tag{10.16}$$

mit

$$\alpha_i := \mu_i - \mu_1 \text{ für } i = 2, \dots, p, \tag{10.17}$$

die Datenverteilungsform

$$\begin{aligned} v_{1j} &\sim N(\mu_0, \sigma^2) && \text{u.i.v. für } j = 1, \dots, n_1 \text{ mit } \mu_0 \in \mathbb{R}, \sigma^2 > 0 \\ v_{ij} &\sim N(\mu_0 + \alpha_i, \sigma^2) && \text{u.v. für } i = 2, \dots, p, j = 1, \dots, n_i \text{ mit } \alpha_i \in \mathbb{R}, \sigma^2 > 0 \end{aligned} \tag{10.18}$$

und die Designmatrixform

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n), n := \sum_{i=1}^p n_i \tag{10.19}$$

$$v := \begin{pmatrix} v_{11} \\ \vdots \\ v_{1n_1} \\ v_{21} \\ \vdots \\ v_{2n_2} \\ \vdots \\ v_{p1} \\ \vdots \\ v_{pn_p} \end{pmatrix} \in \mathbb{R}^n, X := \begin{pmatrix} 1_{n_1} & 0_{n_1} & \dots & 0_{n_1} \\ 1_{n_2} & 1_{n_2} & \dots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1_{n_p} & 0_{n_p} & \dots & 1_{n_p} \end{pmatrix}, \in \mathbb{R}^{n \times p}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \text{ und } \sigma^2 > 0.$$

•

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 $n_i := 4$ und $p = 3$ für $i = 1, \dots, p$, also $n = 12$. Für die Erwartungswertparameterdarstellung gilt dann

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{12}, \sigma^2 I_{12}) \tag{10.20}$$

mit

$$v := \begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \\ v_{21} \\ v_{22} \\ v_{23} \\ v_{31} \\ v_{32} \\ v_{33} \\ v_{41} \\ v_{42} \\ v_{43} \end{pmatrix}, X := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{12 \times 3}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (10.21)$$

Für die Effektparameterdarstellung dagegen gilt

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{12}, \sigma^2 I_{12}) \quad (10.22)$$

mit

$$v := \begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \\ v_{21} \\ v_{22} \\ v_{23} \\ v_{31} \\ v_{32} \\ v_{33} \\ v_{41} \\ v_{42} \\ v_{43} \end{pmatrix}, X := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{12 \times 3}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (10.23)$$

Folgender **R** Code demonstriert die Realisierung von Daten in einem EVA-Szenario mit eben dieser Effektparameterdarstellung.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 m = 4 # Anzahl von Datenpunkten der iten Gruppe
4 p = 3 # Anzahl Gruppen
5 n = p*m # Gesamtanzahl Datenpunkte
6 Xt = cbind( # Designmatrix
7   matrix(1,nrow = n, ncol = 1),
8   kronecker(diag(p), matrix(1,nrow = m,ncol = 1)))
9 X = Xt[,-2]
10 I_n = diag(n) # n x n Einheitsmatrix
11 beta = matrix(c(10,-3,-12), nrow = p) # \beta = (\mu_0, \alpha_2, \alpha_3, \alpha_4)
12 sigsq = 14 # \sigma^2
13 y = mvrnorm(1, X %*% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs
14 print(X) # Ausgabe der Designmatrix

```

```

      [,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

```

10.3 Modellschätzung

Wir betrachten nun die Betaparameterschätzung in der Effektparameterdarstellung des EVA-Modells mit Referenzgruppe. Entsprechend der Interpretation der Betaparameterkomponenten werden dabei μ_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 *gepoolten Stichprobenvarianz* ergibt, wollen wir hier nicht weiter eingehen.

Theorem 10.2 (Betaparameterschätzung im EVA-Modell). *Gegeben sei die Designmatrixform des EVA in Effektdarstellung mit Referenzgruppe Dann ergibt sich für den Betaparameterschätzer*

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 - \bar{v}_1 \\ \vdots \\ \bar{v}_p - \bar{v}_1 \end{pmatrix} \tag{10.24}$$

wobei

$$\bar{v}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \tag{10.25}$$

das Stichprobenmittel der *iten* Gruppe bezeichnet.

Beweis. Wir halten zunächst fest, dass

$$X^T X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix}$$

$$= \begin{pmatrix} n & n_2 & n_3 & \dots & n_p \\ n_2 & n_2 & 0 & \dots & 0 \\ n_3 & 0 & n_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_p & 0 & 0 & \dots & n_p \end{pmatrix}.$$

Die Inverse von $X^T X$ ist

$$(X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & \dots & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \dots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_1} & \frac{1}{n_1} & \dots & \frac{n_1+n_p}{n_1 n_p} \end{pmatrix}. \tag{10.26}$$

So gilt zum Beispiel für $p = 3$, dass

$$X^T X = \begin{pmatrix} n & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{pmatrix} \text{ und } (X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{1}{n_1} & \frac{n_1+n_3}{n_1 n_3} \end{pmatrix}. \tag{10.27}$$

Wir halten weiterhin fest, dass

$$\begin{aligned}
 X^T v &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} v_{11} \\ \vdots \\ v_{1n_1} \\ v_{21} \\ \vdots \\ v_{2n_2} \\ \vdots \\ v_{p1} \\ \vdots \\ v_{pn_p} \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} \\ \sum_{j=1}^{n_2} v_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} v_{pj} \end{pmatrix}.
 \end{aligned} \tag{10.28}$$

Es ergibt sich also

$$\hat{\beta} = (X^T X)^{-1} X^T v = \begin{pmatrix} \frac{1}{n_1} & -\frac{1}{n_1} & \dots & -\frac{1}{n_1} \\ -\frac{1}{n_1} & \frac{n_1+n_2}{n_1 n_2} & \dots & \frac{1}{n_1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_1} & \frac{1}{n_1} & \dots & \frac{n_1+n_p}{n_1 n_p} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} \\ \sum_{j=1}^{n_2} v_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} v_{pj} \end{pmatrix}. \tag{10.29}$$

Für die erste Komponente von $\hat{\beta}$ ergibt sich damit

$$\begin{aligned}
 \hat{\beta}_0 &= \frac{1}{n_1} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} - \frac{1}{n_1} \sum_{j=1}^{n_2} v_{2j} - \dots - \frac{1}{n_1} \sum_{j=1}^{n_p} v_{pj} \\
 &= \frac{1}{n_1} \left(\left(\sum_{j=1}^{n_1} v_{1j} + \sum_{j=1}^{n_2} v_{2j} + \dots + \sum_{j=1}^{n_p} v_{pj} \right) - \sum_{j=1}^{n_2} v_{2j} - \dots - \sum_{j=1}^{n_p} v_{pj} \right) \\
 &= \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\
 &= \bar{v}_1.
 \end{aligned} \tag{10.30}$$

Für die zweite Komponente von $\hat{\beta}$ und analog für alle weiteren ergibt sich

$$\begin{aligned}
 \hat{\beta}_2 &= -\frac{1}{n_1} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} + \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} v_{2j} + \frac{1}{n_1} \sum_{j=1}^{n_3} v_{3j} + \dots + \frac{1}{n_1} \sum_{j=1}^{n_p} v_{pj} \\
 &= \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{1}{n_1} \left(\left(\sum_{j=1}^{n_1} v_{1j} + \sum_{j=1}^{n_2} v_{2j} + \dots + \sum_{j=1}^{n_p} v_{pj} \right) - \sum_{j=1}^{n_3} v_{3j} - \dots - \sum_{j=1}^{n_p} v_{pj} \right) \\
 &= \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} - \frac{1}{n_1} \sum_{j=1}^{n_2} v_{2j} \\
 &= \frac{n_1+n_2}{n_1 n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{n_2}{n_1 n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\
 &= \frac{n_1}{n_1 n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\
 &= \frac{1}{n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} = \bar{v}_2 - \bar{v}_1.
 \end{aligned} \tag{10.31}$$

□

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 evaluiert der Code neben dem Varianzparameterschätzer auch die gepoolte Stichprobenvarianz sowie die Gesamtstichprobenvarianz, die sich deutlich unterscheiden.

```

1 # Dateneinlesen
2 fname = file.path("../data/410-Einfaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # Datengruppen
6 y_1 = D$dBDI[D$COND == "F2F"] # BDI Differenzwerte in der F2F Gruppe
7 y_2 = D$dBDI[D$COND == "ONL"] # BDI Differenzwerte in der ONL Gruppe
8 y_3 = D$dBDI[D$COND == "WLC"] # BDI Differenzwerte in der ONL Gruppe
9
10 # Modellformulierung
11 p = 3 # drei Gruppen
12 m = length(y_1) # balancierteres Design mit n_i = 40
13 n = p*m # Datenvektordimension
14 y = matrix(c(y_1, y_2, y_3), nrow = n) # Datenvektor
15 Xt = cbind( # Designmatrix
16   matrix(1, nrow = n, ncol = 1),
17   kronecker(diag(p), matrix(1, nrow = m, ncol = 1)))
18 X = Xt[,-2]
19
20 # Modellschätzung
21 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
22 eps_hat = y - X %*% beta_hat # Residuenvektor
23 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
24 s_sqr_123 = ((m-1)*var(y_1) + # gepoolte Stichprobenvarianz
25   (m-1)*var(y_2) +
26   (m-1)*var(y_3)) / (m+m+m-p)
27
28 # Ausgabe
29 cat("hat{beta} : ", beta_hat,
30   "\nbar{y}_1,bar{y}_2,bar{y}_3 : ", c(mean(y_1),mean(y_2),mean(y_3)),
31   "\nbar{y}_1,bar{y}_2-bar{y}_1,bar{y}_3-bar{y}_1 : ", c(mean(y_1),mean(y_2)-mean(y_1),mean(y_3)-mean(y_1)),
32   "\nhat{sigsqr} : ", sigsq_hat,
33   "\ns_123^2 : ", s_sqr_123,
34   "\ns_y^2 : ", var(y))

```

```

hat{beta} : 9.833333 -2.75 -11
bar{y}_1,bar{y}_2,bar{y}_3 : 9.833333 7.083333 -1.166667
bar{y}_1,bar{y}_2-bar{y}_1,bar{y}_3-bar{y}_1 : 9.833333 -2.75 -11
hat{sigsqr} : 6.916667
s_123^2 : 6.916667
s_y^2 : 28.99286

```

10.4 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 gleich Null sind, formal

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \mid \alpha_i = 0 \text{ für } i = 2, \dots, p \right\} = \mathbb{R} \times \{0_{p-1}\} \quad (10.32)$$

und

$$\Theta_1 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \mid \alpha_i \neq 0 \text{ für mindestens ein } i = 2, \dots, p \right\} = \mathbb{R}^p \setminus \Theta_0. \quad (10.33)$$

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 η^2 (Eta-Quadrat) auch ein zu dem aus Kapitel 2.2 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.

10.4.1 Quadratsummenzerlegung und Bestimmtheitsmaß η^2

Die Variabilität der Daten eines EVA-Szenarios lässt sich wie in folgendem Theorem dargestellt im Sinne einer *Quadratsummenzerlegung* schreiben.

Theorem 10.3 (Quadratsummenzerlegung bei einfaktorieller Varianzanalyse). *Für $i = 1, \dots, p$ und $j = 1, \dots, n_i$ sei v_{ij} die j te Datenvariable in der i ten Gruppe eines EVA-Szenarios. Weiterhin seien mit $n := \sum_{i=1}^p n_i$*

$$\bar{v} := \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} \text{ und } \bar{v}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \quad (10.34)$$

das Gesamtstichprobenmittel und i te Stichprobenmittel, respektive. Schließlich seien

- die Total Sum of Squares definiert als

$$SQT := \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v})^2 \quad (10.35)$$

- die Between Sum of Squares definiert als

$$SQB := \sum_{i=1}^p n_i (\bar{v}_i - \bar{v})^2 \quad (10.36)$$

und die Within Sum of Squares definiert als

$$SQW := \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 \quad (10.37)$$

Dann gilt

$$SQT = SQB + SQW. \quad (10.38)$$

Beweis. Es gilt

$$\begin{aligned}
\text{SQT} &= \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v})^2 \\
&= \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i + \bar{v}_i - \bar{v})^2 \\
&= \sum_{i=1}^p \sum_{j=1}^{n_i} ((v_{ij} - \bar{v}_i) + (\bar{v}_i - \bar{v}))^2 \\
&= \sum_{i=1}^p \sum_{j=1}^{n_i} ((v_{ij} - \bar{v}_i)^2 + 2(v_{ij} - \bar{v}_i)(\bar{v}_i - \bar{v}) + (\bar{v}_i - \bar{v})^2) \\
&= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + \sum_{j=1}^{n_i} 2(v_{ij} - \bar{v}_i)(\bar{v}_i - \bar{v}) + \sum_{j=1}^{n_i} (\bar{v}_i - \bar{v})^2 \right)
\end{aligned} \tag{10.39}$$

und weiter

$$\begin{aligned}
\text{SQT} &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + 2(\bar{v}_i - \bar{v}) \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i) + n_i(\bar{v}_i - \bar{v})^2 \right) \\
&= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + 2(\bar{v}_i - \bar{v}) \sum_{j=1}^{n_i} \left(v_{ij} - \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \right) + n_i(\bar{v}_i - \bar{v})^2 \right) \\
&= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + 2(\bar{v}_i - \bar{v}) \left(\sum_{j=1}^{n_i} v_{ij} - \sum_{j=1}^{n_i} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \right) \right) + n_i(\bar{v}_i - \bar{v})^2 \right) \\
&= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + 2(\bar{v}_i - \bar{v}) \left(\sum_{j=1}^{n_i} v_{ij} - \frac{n_i}{n_i} \sum_{j=1}^{n_i} v_{ij} \right) + n_i(\bar{v}_i - \bar{v})^2 \right)
\end{aligned} \tag{10.40}$$

und weiter

$$\begin{aligned}
\text{SQT} &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + 2(\bar{v}_i - \bar{v}) \left(\sum_{j=1}^{n_i} v_{ij} - \sum_{j=1}^{n_i} v_{ij} \right) + n_i(\bar{v}_i - \bar{v})^2 \right) \\
&= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + n_i(\bar{v}_i - \bar{v})^2 \right) \\
&= \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 + \sum_{i=1}^p n_i(\bar{v}_i - \bar{v})^2 \\
&= \text{SQW} + \text{SQB}
\end{aligned} \tag{10.41}$$

und damit direkt

$$\text{SQT} = \text{SQB} + \text{SQW}. \tag{10.42}$$

□

Die Quadratsummenzerlegung im Szenario der EVA ist offenbar analog zur Quadratsummenzerlegung bei einer Ausgleichsgerade (vgl. Kapitel 2.2). 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 wiederum wie folgt:

- SQT repräsentiert die Gesamtvariabilität der Daten v_{ij} um das Gesamtstichprobenmittel \bar{v} .
- 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 Explained Sum of Squares (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 Residual Sum of Squares der Quadratsummenzerlegung bei Ausgleichsgerade.

In der Zusammenschau quantifiziert SQB also die Stärke der Unterschiede zwischen den Faktorleveln. Das Effektstärkenmaß η^2 und die F-Statistik der EVA setzen die SQB nun jeweils in ein anderes Verhältnis: η^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 in das Verhältnis mit der nicht erklärten Datenvariabilität nach Subtraktion dieses Einflusses.

Wir definieren zunächst das Effektstärkemaß η^2 .

Definition 10.3 (Effektstärkenmaß η^2). Für ein EVA-Szenario seien die Between Sum of Squares SQB und die Total Sum of Squares SQT definiert wie in Theorem 10.3. Dann ist das *Effektstärkenmaß* η^2 definiert als

$$\eta^2 := \frac{\text{SQB}}{\text{SQT}} \tag{10.43}$$

•

Der Vergleich mit Definition 2.5 zeigt, dass η^2 analog zum Bestimmtheitsmaß R^2 definiert ist. Wie oben beschrieben gibt η^2 den Anteil der Datenvariabilität zwischen den Faktorleveln an der Gesamtdatenvariabilität. Schließlich folgt mit

$$\text{SQT} = \text{SQB} + \text{SQW}. \tag{10.44}$$

analog zu R^2 sofort, dass für $\text{SQT} \neq 0$ gilt, dass $\eta^2 \in [0, 1]$, weil einerseits

$$\text{SQB} = 0 \Rightarrow \text{SQT} = \text{SQW} \text{ und } \eta^2 = 0 \tag{10.45}$$

und andererseits

$$\text{SQW} = 0 \Rightarrow \text{SQT} = \text{SQB} \text{ und } \eta^2 = 1. \tag{10.46}$$

10.4.2 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 und Anzahl der Betaparameter, respektive. Es gilt folgendes Theorem.

Theorem 10.4 (F-Statistik der einfaktoriellen Varianzanalyse). *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{10.47}$$

die Designmatrixform der Effektdarstellung mit Referenzgruppe des EVA-Modells. Weiterhin sei dieses Modell im Sinne von Definition 8.2 partitioniert mit $p_0 := 1$ und $p_1 := p - 1$. Schließlich seien

- die Mean Between Sum of Squares definiert als

$$MSB := \frac{SQB}{p - 1} \tag{10.48}$$

- und die Mean Within Sum of Squares definiert als

$$MSW := \frac{SQW}{n - p}, \tag{10.49}$$

wobei $p - 1$ auch als Between Freiheitsgrade und $n - p$ auch als Within Freiheitsgrade bezeichnet werden. Dann gilt mit der Definition der F-Statistik in Definition 8.3, dass

$$F = \frac{MSB}{MSW}. \tag{10.50}$$

Beweis. Wir halten zunächst fest, dass für den Betaparameterschätzer des reduzierten Modells gilt, dass

$$\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T v = (1_n^T 1_n)^{-1} 1_n^T v = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} = \bar{v}. \tag{10.51}$$

Weiterhin ergibt sich

$$\hat{\varepsilon}_0^T \hat{\varepsilon}_0 = (v - X_0 \hat{\beta}_0)^T (v - X_0 \hat{\beta}_0) = (v - 1_n \bar{v})^T (v - 1_n \bar{v}) = \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v})^2 = \text{SQT}. \tag{10.52}$$

Der Betaparameterschätzer des vollständigen Modells ergibt sich wie oben gesehen zu

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_m \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} v_{2j} - \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \\ \vdots \\ \frac{1}{n_m} \sum_{j=1}^{n_m} v_{mj} - \frac{1}{n_1} \sum_{j=1}^{n_1} v_{1j} \end{pmatrix} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 - \bar{v}_1 \\ \vdots \\ \bar{v}_p - \bar{v}_1 \end{pmatrix}, \tag{10.53}$$

so dass

$$\begin{aligned} \hat{\varepsilon}^T \hat{\varepsilon} &= (v - X\hat{\beta})^T (v - X\hat{\beta}) \\ &= \begin{pmatrix} \begin{pmatrix} v_{11} \\ \vdots \\ v_{1n_1} \\ v_{21} \\ \vdots \\ v_{2n_2} \\ \vdots \\ v_{p1} \\ \vdots \\ v_{pn_p} \end{pmatrix} - \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 - \bar{v}_1 \\ \vdots \\ \bar{v}_p - \bar{v}_1 \end{pmatrix} \end{pmatrix}^T (v - X\hat{\beta}) \end{aligned} \tag{10.54}$$

und weiter

$$\begin{aligned}
 \hat{\varepsilon}^T \hat{\varepsilon} &= \begin{pmatrix} v_{11} - \bar{v}_1 \\ \vdots \\ v_{1n_1} - \bar{v}_1 \\ v_{21} - \bar{v}_1 - \bar{v}_2 + \bar{v}_1 \\ \vdots \\ v_{2n_2} - \bar{v}_1 - \bar{v}_2 + \bar{v}_1 \\ \vdots \\ v_{p1} - \bar{v}_1 - \bar{v}_p + \bar{v}_1 \\ \vdots \\ v_{pn_p} - \bar{v}_1 - \bar{v}_p + \bar{v}_1 \end{pmatrix}^T \begin{pmatrix} v_{11} - \bar{v}_1 \\ \vdots \\ v_{1n_1} - \bar{v}_1 \\ v_{21} - \bar{v}_1 - \bar{v}_2 + \bar{v}_1 \\ \vdots \\ v_{2n_2} - \bar{v}_1 - \bar{v}_2 + \bar{v}_1 \\ \vdots \\ v_{p1} - \bar{v}_1 - \bar{v}_p + \bar{v}_1 \\ \vdots \\ v_{pn_p} - \bar{v}_1 - \bar{v}_p + \bar{v}_1 \end{pmatrix} = \begin{pmatrix} v_{11} - \bar{v}_1 \\ \vdots \\ v_{1n_1} - \bar{v}_1 \\ v_{21} - \bar{v}_2 \\ \vdots \\ v_{2n_2} - \bar{v}_2 \\ \vdots \\ v_{p1} - \bar{v}_p \\ \vdots \\ v_{pn_p} - \bar{v}_p \end{pmatrix}^T \begin{pmatrix} v_{11} - \bar{v}_1 \\ \vdots \\ v_{1n_1} - \bar{v}_1 \\ v_{21} - \bar{v}_2 \\ \vdots \\ v_{2n_2} - \bar{v}_2 \\ \vdots \\ v_{p1} - \bar{v}_p \\ \vdots \\ v_{pn_p} - \bar{v}_p \end{pmatrix} \\
 &= \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)^2 \\
 &= \text{SQW}.
 \end{aligned} \tag{10.55}$$

Mit dem Theorem zur Quadratsummenzerlegung bei einfaktorieller Varianzanalyse

$$\text{SQT} = \text{SQB} + \text{SQW} \Leftrightarrow \text{SQB} = \text{SQT} - \text{SQW} \tag{10.56}$$

folgt sofort, dass

$$\begin{aligned}
 \text{SQB} &= \text{SQT} - \text{SQW} \\
 &= \hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}.
 \end{aligned} \tag{10.57}$$

Dann aber folgt auch direkt, dass

$$\begin{aligned}
 \frac{\text{MSB}}{\text{MSW}} &= \frac{\frac{\text{SQB}}{p-1}}{\frac{\text{SQW}}{n-p}} \\
 &= \frac{\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon}}{\frac{p-1}{n-p}} \\
 &= F
 \end{aligned} \tag{10.58}$$

□

Gegeben die sehr ähnlichen Definitionen des Effektstärkemaßes η^2 und der F-Statistik der EVA ist folgendes Resultat naheliegend.

Theorem 10.5 (Effektstärkenmaß η^2 und F-Teststatistik). *Für ein EVA-Szenario mit p Gruppen und Gesamtdatenpunktzahl n seien das Effektstärkenmaß η^2 und die F-Statistik der einfaktoriellen Varianzanalyse. Dann gilt*

$$\eta^2 = \frac{F(p-1)}{F(p-1) + (n-p)} \tag{10.59}$$

Beweis. Wir halten zunächst fest, dass

$$F = \frac{\text{SQB}}{\text{SQW}} \cdot \frac{n-p}{p-1} \Leftrightarrow \text{SQB} = \frac{p-1}{n-p} \cdot \text{SQW} \cdot F \tag{10.60}$$

Damit folgt dann

$$\begin{aligned}
 \eta^2 &= \frac{\text{SQB}}{\text{SQT}} = \frac{\text{SQB}}{\text{SQB} + \text{SQW}} = \frac{\frac{p-1}{n-p} \cdot \text{SQW} \cdot F}{\frac{p-1}{n-p} \cdot \text{SQW} \cdot F + \text{SQW}} = \frac{\frac{F(p-1)}{n-p} \cdot \text{SQW}}{\frac{F(p-1)}{n-p} \cdot \text{SQW} + \text{SQW}} \\
 &= \frac{\frac{F(p-1)}{n-p} \cdot \text{SQW}}{\left(\frac{F(p-1)}{n-p} + 1\right) \cdot \text{SQW}} \\
 &= \frac{\frac{F(p-1)}{n-p}}{\frac{F(p-1)}{n-p} + \frac{n-p}{n-p}} \\
 &= \frac{\frac{F(p-1)}{n-p}}{\frac{F(p-1) + (n-p)}{n-p}} \\
 &= \frac{F(p-1)}{F(p-1) + (n-p)}
 \end{aligned} \tag{10.61}$$

□

Intuitiv ist dabei das Verhältnis von η^2 und F-Statistik analog zum Verhältnis von Cohen's d und der T-Statistik bei Einstich- und Zweistichproben-T-Tests. Insbesondere ist die gleichzeitige Angabe von η^2 und der F-Statistik bei bekannten Gruppengrößen redundant.

10.4.3 F-Test der einfaktoriellen Varianzanalyse

Wir wollen nun den Gebrauch der F-Statistik zur Durchführung eines Tests der Nullhypothese (Gleichung 10.32) 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 EVA-Modell in Effektparameterdarstellung mit Referenzgruppe sowie die Null- und Alternativhypothesen

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \in \mathbb{R}^p \mid \alpha_i = 0 \text{ für } i = 2, \dots, p \right\} \text{ und } \Theta_1 := \mathbb{R}^p \setminus \Theta_0. \tag{10.62}$$

Weiterhin sei die F-Teststatistik definiert durch

$$F = \frac{\text{MSB}}{\text{MSW}} \tag{10.63}$$

mit der Mean Sum of Squares Between MSB und der Mean Sum of Squares Within MSW definiert wie in Theorem 10.4. Dann ist der *F-Test der einfaktoriellen Varianzanalyse* definiert als der kritische wertbasierte Test

$$\phi(v) := 1_{\{F \geq k\}} := \begin{cases} 1 & F \geq k \\ 0 & F < k \end{cases}. \tag{10.64}$$

•

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). ϕ sei der F-Test der einfaktoriellen Varianzanalyse. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \varphi^{-1}(1 - \alpha_0; p - 1, n - p), \tag{10.65}$$

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

$$q : \mathbb{R} \rightarrow [0, 1], \beta \mapsto q_\phi(\beta) := \mathbb{P}_\beta(\phi = 1). \tag{10.66}$$

Wir haben in Einheit Kapitel 8 gesehen, dass die F-Statistik für $p_1 = p - 1$ nach einer nichtzentralen f -Verteilung verteilt ist,

$$F \sim f(\delta, p - 1, n - p). \tag{10.67}$$

Weiterhin ist der Ablehnungsbereich des hier betrachteten Tests gegeben als $[k, \infty[$. Für die funktionale Form der Testgütefunktion ergibt sich also

$$\begin{aligned} \mathbb{P}_\beta(\phi = 1) &= \mathbb{P}_\beta(F \in [k, \infty[) \\ &= \mathbb{P}_\beta(F \geq k) \\ &= 1 - \mathbb{P}_\beta(F \leq k) \\ &= 1 - \varphi(k; \delta, p - 1, n - p), \end{aligned} \tag{10.68}$$

wobei $\varphi(k; \delta, p - 1, n - p)$ den Wert der KVF der nichtzentralen f -Verteilung an der Stelle k und mit Nichtzentralitätsparameter δ sowie Freiheitsgradparametern $p - 1$ und $n - p$ bezeichnet (vgl. Kapitel 8). Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich gelten, dass

$$q_\phi(\beta) \leq \alpha_0 \text{ für alle } \beta \in \Theta_0 \text{ mit } \Theta_0 = \mathbb{R} \times \{0_{p-1}\}. \tag{10.69}$$

Mit der Form des Nichtzentralitätsparameters gegeben durch (vgl. Theorem 8.3)

$$\delta = \frac{1}{\sigma^2} c^T \beta (c^T (X^T X)^{-1} c)^{-1} c^T \beta \tag{10.70}$$

folgt mit $\beta \in \Theta_0$ aus

$$c = \begin{pmatrix} 0 \\ 1_{p-1} \end{pmatrix} \in \mathbb{R}^p \text{ und } \beta = \begin{pmatrix} \mu_0 \\ 0_{p-1} \end{pmatrix} \in \mathbb{R}^p \tag{10.71}$$

dann aber $\delta = 0$ und somit

$$q_\phi(\beta) = 1 - \varphi(k; p - 1, n - p) \text{ für alle } \beta \in \Theta_0. \tag{10.72}$$

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. Der Testumfang des betrachteten Tests ergibt sich als

$$\alpha = \max_{\beta \in \Theta_0} q_\phi(\beta) = 1 - \varphi(k; p - 1, n - p), \tag{10.73}$$

da $q_\phi(\beta)$ für $\beta \in \Theta_0$ nicht von μ_0 abhängt. Wir müssen also lediglich zeigen, dass die Wahl von k_{α_0} wie im Theorem garantiert, dass ϕ den Testumfang α_0 hat. Sei also $k := k_{\alpha_0}$. Dann gilt für alle $\beta \in \Theta_0$

$$q_\phi(\beta) = 1 - \varphi(\varphi^{-1}(1 - \alpha_0; p - 1, n - p); p - 1, n - p) = 1 - 1 + \alpha_0 = \alpha_0 \tag{10.74}$$

und damit ist alles gezeigt.

□

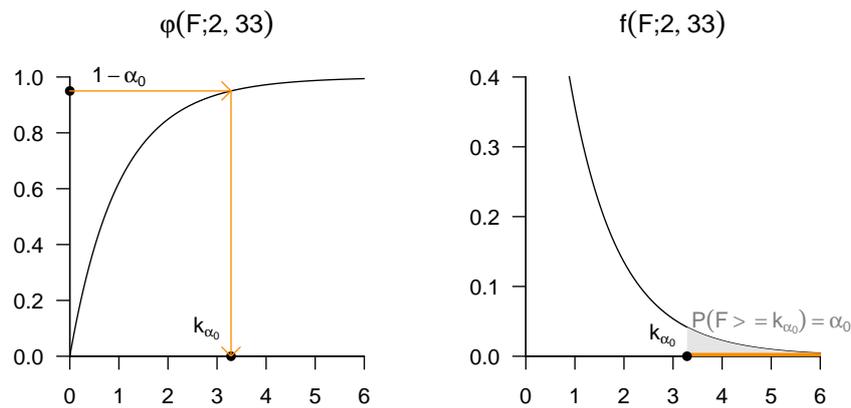


Abbildung 10.2 Exemplarischer kritischer Wert und Ablehnungsbereich bei einem F-Test der einfaktoriellen Varianzanalyse.

In Abbildung 10.2 visualisieren wir die Wahl des kritischen Wertes k_{α_0} zur Kontrolle des Testumfangs mithilfe der KVF der f -Verteilung sowie den resultierenden Ablehnungsbereich für $\alpha_0 := 0.05$, einem Faktor mit drei Leveln $p = 3$ und einem balancierten Design mit Gruppenumfang $m = 12$, also $n = 3 \cdot 12 = 36$. 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.

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*

$$p\text{-Wert} = \mathbb{P}(F \geq f) = 1 - \varphi(f; p - 1, n - p). \quad (10.75)$$

Beweis. Nach Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegenden Wert der Teststatistik ablehnen würde. Bei $F = f$ würde H_0 für jedes α_0 mit $f \geq \psi^{-1}(1 - \alpha_0; p - 1, n - p)$ abgelehnt werden. Für ein solches α_0 gilt aber

$$\alpha_0 \geq \mathbb{P}(F \geq f), \quad (10.76)$$

denn

$$\begin{aligned} f &\geq \psi^{-1}(1 - \alpha_0; p - 1, n - p) \\ \Leftrightarrow \psi(f; p - 1, n - p) &\geq \psi(\psi^{-1}(1 - \alpha_0; p - 1, n - p), p - 1, n - p) \\ \Leftrightarrow \psi(f; p - 1, n - p) &\geq 1 - \alpha_0 \\ \Leftrightarrow \mathbb{P}(F \leq f) &\geq 1 - \alpha_0 \\ \Leftrightarrow \alpha_0 &\geq 1 - \mathbb{P}(F \leq f) \\ \Leftrightarrow \alpha_0 &\geq \mathbb{P}(F \geq f) \end{aligned} \quad (10.77)$$

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

$$p\text{-Wert} = \mathbb{P}(F \geq f) = 1 - \varphi(f; p - 1, n - p). \quad (10.78)$$

□

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

$$y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2}, \dots, y_{p1}, \dots, y_{pn_p} \quad (10.79)$$

Realisationen von

$$v_{1j} \sim N(\mu_0, \sigma^2) \quad (10.80)$$

und

$$v_{ij} \sim N(\mu_0 + \alpha_i, \sigma^2) \quad (10.81)$$

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 10.32).

Dazu wählt man zunächst ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05, p = 3, m = 12, i = 1, 2, 3$ und somit $n = 36$, dass $k_{\alpha_0} = \varphi^{-1}(1 - 0.05; 2, 33) \approx 3.28$ ist. Man nutzt dann den vorliegenden Datensatz zur Berechnung der MSB und MSW und bestimmt damit den realisierten Wert der F-Teststatistik f . Wenn f größer gleich k_{α_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.

10.4.4 Anwendungsbeispiel

Folgender **R** Code implementiert das obige praktische Vorgehen für den Beispieldatensatz. Im vorliegenden Fall würde die Nullhypothese

$$\Theta_0 := \left\{ \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \in \mathbb{R}^3 \mid \alpha_i = 0 \text{ für } i = 2, 3 \right\} \quad (10.82)$$

verworfen werden.

```

1 # Dateneinlesen
2 fname = file.path("../data/410-Einfaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # Datengruppen
6 y_1 = D$dBDI[D$COND == "F2F"] # BDI Differenzwerte in der F2F Gruppe
7 y_2 = D$dBDI[D$COND == "ONL"] # BDI Differenzwerte in der ONL Gruppe
8 y_3 = D$dBDI[D$COND == "WLC"] # BDI Differenzwerte in der ONL Gruppe
9
10 # Modellformulierung
11 p = 3 # drei Gruppen
12 m = length(y_1) # balanciertes Design mit n_i = 40
13 n = p*m # Datenvektordimension
14 y = matrix(c(y_1, y_2, y_3), nrow = n) # Datenvektor
15 Xt = cbind( # Designmatrix vollständiges Modell
16   matrix(1, nrow = n, ncol = 1),
17   kronecker(diag(p),
18     matrix(1, nrow = m, ncol = 1)))

```

```

19 X      = Xt[,-2]
20 X_0    = X[,1]                                # Designmatrix reduziertes Modell
21
22 # F-Teststatistikevaluation
23 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y    # Betaparameterschätzer vollständiges Modell
24 beta_hat_1 = solve(t(X_0) %*% X_0) %*% t(X_0) %*% y # Betaparameterschätzer reduziertes Modell
25 eps_hat   = y - X %*% beta_hat                # Residuenvektor vollständiges Modell
26 eps_hat_1 = y - X_0 %*% beta_hat_1            # Residuenvektor reduziertes Modell
27 SQT       = t(eps_hat_1) %*% eps_hat_1        # Sum of Squares Total
28 SQW       = t(eps_hat) %*% eps_hat            # Sum of Squares Within
29 SQB       = SQT - SQW                         # Sum of Squares Between
30 DFB       = p - 1                             # Between Degrees of Freedom
31 DFW       = n - p                             # Within Degrees of Freedom
32 DFB       = p - 1                             # Between Degrees of Freedom
33 MSB       = SQB/DFB                           # Mean Sum of Squares Between
34 MSW       = SQW/DFW                           # Mean Sum of Squares Within
35 Eff       = MSB/MSW                           # F-Teststatistik
36 p         = 1 - pf(Eff, p-1, n-p)             # p-Wert
37
38 # Ausgabe
39 cat("DFB :", DFB,
40     "\nDFW :", DFW,
41     "\nSQB :", SQB,
42     "\nSQW :", SQW,
43     "\nMSB :", MSB,
44     "\nMSW :", MSW,
45     "\nF   :", Eff,
46     "\np   :", paste(p))

```

```

DFB : 2
DFW : 33
SQB : 786.5
SQW : 228.25
MSB : 393.25
MSW : 6.916667
F   : 56.85542
p   : 2.03634886730697e-11

```

Folgender **R** Code demonstriert die Durchführung und Dokumentation der gleichen Analyse mithilfe der **R** Funktion `aov()`.

```

1 # Dateneinlesen
2 fname = file.path("../data/410-Einfaktorielle-Varianzanalyse.csv")
3 D      = read.table(fname, sep = ",", header = TRUE)
4
5 # Benutzung von R's aov Funktion und Ausgabe
6 res.aov = aov(D$dBDI ~ D$COND, data = D)
7 summary(res.aov)

```

```

          Df Sum Sq Mean Sq F value    Pr(>F)
D$COND    2  786.5    393.2    56.85 2.04e-11 ***
Residuals 33  228.2     6.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

10.5 Literaturhinweise

Die Popularität varianzanalytischer Verfahren wird im Allgemeinen auf Fisher (1925b) und Fisher (1935) zurückgeführt. Everitt und Howell (2005) und Stigler (1986) geben einen kurzen und einen ausführlichen historischen Überblick, respektive.

10.6 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 Reparametrisierung des EVA-Modells
5. Welche Bedeutung haben $\mu_0, \alpha_2, \dots, \alpha_p$ 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 Deskriptivstatistiken werden die Parameter $\mu_0, \alpha_2, \dots, \alpha_p$ geschätzt?
12. Geben Sie das Theorem zur Quadratsummenzerlegung bei einfaktorieller Varianzanalyse wieder.
13. Erläutern Sie die Begriffe der Total, Between, und Within Sum of Squares der EVA.
14. Geben Sie die Definition des Effektstärkenmaßes η^2 an.
15. Wann nimmt das Effektstärkenmaß η^2 der EVA seinen Minimalwert an und wie lautet dieser?
16. Wann nimmt das Effektstärkenmaß η^2 der EVA seinen Maximalwert an und wie lautet dieser?
17. Geben Sie das Theorem zur F-Teststatistik der EVA wieder.
18. Erläutern Sie die Begriffe Mean Between Sum of Squares und Mean Within Sum of Squares der EVA.
19. Geben Sie das Theorem zum Zusammenhang von η^2 und F-Teststatistik 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 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

2×2 ZVA	\Rightarrow	Faktor A mit Level 1,2	Faktor B mit Level 1,2
2×2 ZVA	\Rightarrow	Faktor A mit Level 1,2	Faktor B mit Level 1,2,3
2×2 ZVA	\Rightarrow	Faktor A mit Level 1,2,3,4	Faktor B mit Level 1,2
2×2 ZVA	\Rightarrow	Faktor A mit Level 1,2,3	Faktor B mit Level 1

Generell sind 2×2 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 A visualisieren wir das konzeptuelle Design einer 2×2 ZVA, wobei wir die Zellen des Designs hier und im Folgenden mit A1B1, A1B2, A2B1 und A2B2 bezeichnen.

Abbildung 11.1 B 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*.

Haupteffekte und Interaktionen

Intuitiv ist man in ZVA Designs an *Haupteffekten* und *Interaktionen* interessiert, die sich zunächst rein deskriptiv auf das Muster der zellspezifischen Gruppenmittelwerte beziehen.

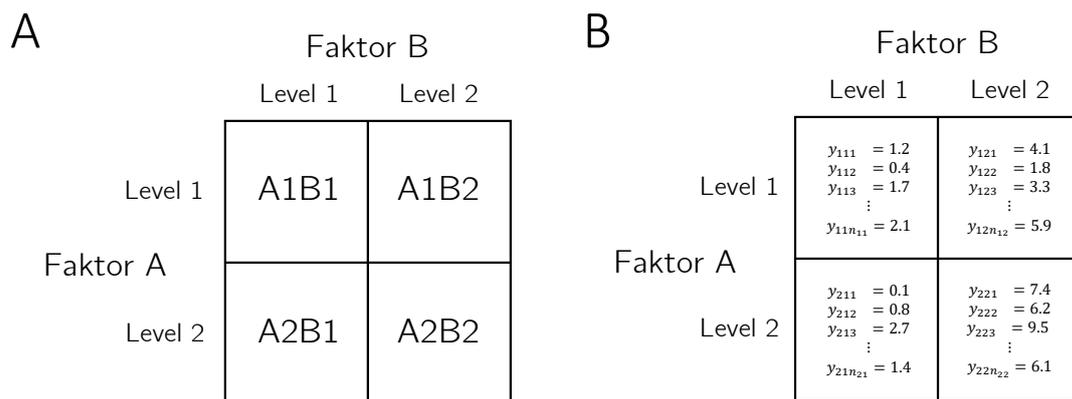


Abbildung 11.1 Konzeptuelles Design und Datennotation einer 2×2 ZVA

Für den Fall einer 2×2 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 frequentitistisch-inferenzstatistische Absicherung dieser Datenmuster ist das Thema dieses Kapitels.

Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel aus dem Bereich der Interventionsevaluation betrachten wir eine BDI-Pre-Postinterventionsdifferenzwertanalyse für je zwei Settings (Face-to-Face und Online) und Varianten (Mindfulness und Exercise) der kognitiven Verhaltenstherapie (CBT, Abbildung 11.2). 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 zwischen Preinterventions und Postinterventions BDI Wert bestimmt worden und mit $dBDI$ bezeichnet.

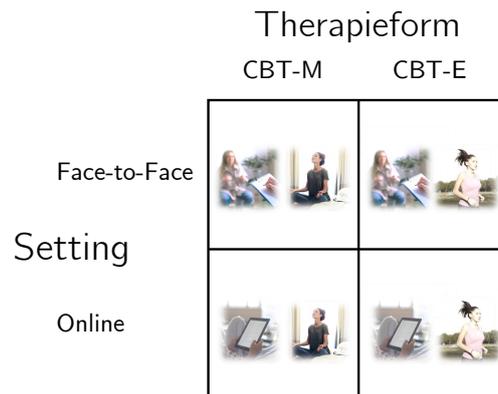


Abbildung 11.2 Konzeptuelles Design des Anwendungsbeispiels

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 Zeilenwert bildet den Pre-Post-Interventions-BDI-Differenzwert der entsprechenden Patient:in ab. Hohe Werte zeigen dabei eine starke Verbesserung, geringe Werte eher eine leichte Verbesserung der Depressionssymptomatik an.

Tabelle 11.1 Pre-Postinterventions-BDI-Differenzwerte Werte über Faktorlevel

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	13
F2F	EXC	10
F2F	EXC	14
F2F	EXC	15
F2F	EXC	14
F2F	EXC	18
F2F	EXC	15
F2F	EXC	13
F2F	EXC	10
F2F	EXC	14
F2F	EXC	14
F2F	EXC	16
ONL	MND	4
ONL	MND	10
ONL	MND	12
ONL	MND	13
ONL	MND	12
ONL	MND	12

Setting	Variant	dBDI
ONL	MND	13
ONL	MND	10
ONL	MND	10
ONL	MND	13
ONL	MND	4
ONL	MND	8
ONL	EXC	15
ONL	EXC	18
ONL	EXC	13
ONL	EXC	16
ONL	EXC	16
ONL	EXC	15
ONL	EXC	12
ONL	EXC	15
ONL	EXC	19
ONL	EXC	12
ONL	EXC	15
ONL	EXC	12

Abbildung 11.3 schließlich zeigt zwei häufig genutzte Visualisierungsformen für Deskriptivstatistiken (hier Gruppenmittelwerte und Standardabweichungen) in 2×2 ZVA Szenarien. Abbildung 11.3 A zeigt die faktorlevelkombinationsspezifischen Gruppenmittelwerte und ihre Standardabweichungen als Balkendiagramm. Abbildung 11.3 B zeigt die gleichen Daten als Liniendiagramm mit Fehlerbalken. Dabei entsprechen die zwei Linien den beiden 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 in Abbildung 11.3 A und Abbildung 11.3 B einmal explizit durch nachzeichnen herzustellen.

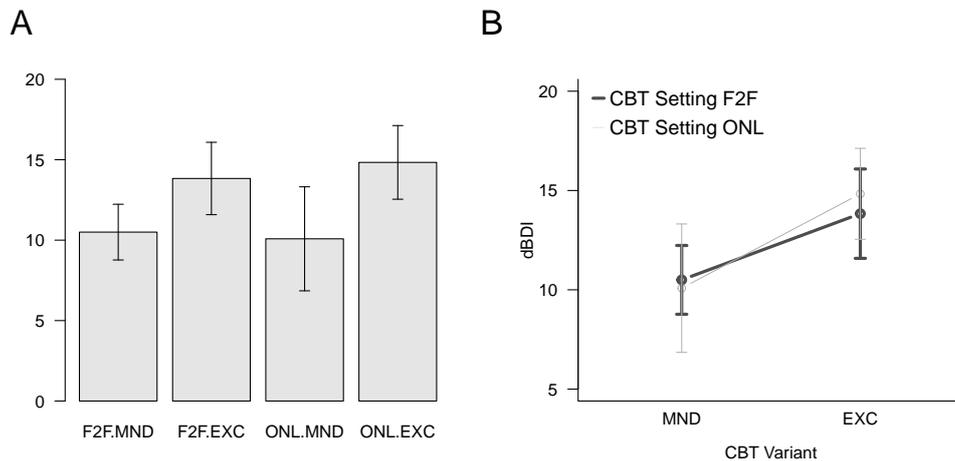


Abbildung 11.3 Visualisierung des Beispieldatensatzes als Balken- und Liniendiagramm

11.2 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.

11.2.1 Modell der additiven ZVA

In Analogie zur EVA möchte man im Modell der additiven ZVA die Gruppenerwartungswerte μ_{ij} mit $i = 1, \dots, I$ für die Level von Faktor A und $j = 1, \dots, J$ für die Level von Faktor B als Summe eines gruppenunspezifischen Erwartungswertes und den Effekten der Level von Faktor A und der Level von Faktor B modellieren.

Wir bezeichnen dabei den gruppenunspezifischen Erwartungswertparameter mit μ_0 , den Effekt von Level i von Faktor A mit α_i und den Effekt von Level j von Faktor B mit β_j (β_j bezeichnet hier also nicht den j ten Eintrag des Betaparametervektors). Dann ergibt sich zum Beispiel für $I := J := 2$

$$\begin{array}{c|c} \mu_{11} := \mu_0 + \alpha_1 + \beta_1 & \mu_{12} := \mu_0 + \alpha_1 + \beta_2 \\ \mu_{21} := \mu_0 + \alpha_2 + \beta_1 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 \end{array}$$

Wie im Falle der EVA ist diese Darstellung der Gruppenerwartungswerte μ_{ij} allerdings überparameterisiert, speziell werden die vier Erwartungswertparameter $\mu_{11}, \mu_{12}, \mu_{21}$ und μ_{22} in obiger Formulierung durch die fünf Parameter $\mu_0, \alpha_1, \alpha_2, \beta_1$ und β_2 dargestellt. Um eine eindeutige Darstellung der μ_{ij} zu gewährleisten, bietet sich auch hier die Restriktion an, den Effekt des ersten Levels jedes Faktors als Null zu definieren

$$\alpha_1 := \beta_1 := 0. \quad (11.1)$$

und damit die Faktorlevelkombination A1B1 als Referenzgruppe zu etablieren. Es ergibt sich somit zum Beispiel für $I := J := 2$

$$\begin{array}{c|c} \mu_{11} := \mu_0 & \mu_{12} := \mu_0 + \beta_2 \\ \mu_{21} := \mu_0 + \alpha_2 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 \end{array}$$

Hier werden nun also die vier Erwartungswertparameter $\mu_{11}, \mu_{12}, \mu_{21}$ und μ_{22} durch nur drei Effektparameter μ_0, α_2 und β_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 2×2 ZVA mit Referenzgruppe analog zur EVA die Interpretationen der Parameter μ_0, α_2, β_2 im Vergleich zum überparameterisierten Fall ohne Referenzgruppe: μ_0 entspricht dem Erwartungswert der Faktorlevelkombination A1B1, α_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A und β_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B. Wir fassen obige Überlegungen in folgender Definition zusammen.

Definition 11.1 (Modell der additiven ZVA mit Referenzgruppe). v_{ijk} mit $i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij}$ sei die Zufallsvariable, die den k ten Datenpunkt zum i ten Level von Faktor A und dem j ten Level von Faktor B in einem ZVA Anwendungsszenario modelliert. Dann hat das *Modell der additiven ZVA mit Referenzgruppe* die strukturelle Form

$$v_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.2)$$

und die Datenverteilungsform

$$v_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.3)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j \text{ für } i = 1, \dots, I, j = 1, \dots, J \text{ mit } \alpha_1 := \beta_1 := 0. \quad (11.4)$$

und $\sigma^2 > 0$.

•

Wir verzichten in Definition 11.2 auf die Angabe einer allgemeinen Designmatrixform, die wir im Folgenden lediglich für den 2×2 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 11.4.

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

$$\begin{aligned} \mu_{11} &= \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = 1 & \mu_{12} &= \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 0 = 1 \\ \mu_{21} &= \mu_0 + \alpha_2 + \beta_1 = 1 + 1 + 0 = 2 & \mu_{22} &= \mu_0 + \alpha_2 + \beta_2 = 1 + 1 + 0 = 2 \end{aligned}$$

Abbildung 11.4 A zeigt Visualisierungen dieses Erwartungswertmusters in der Balkendiagramm und Liniendiagrammform von Abbildung 11.3.

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

$$\begin{aligned} \mu_{11} &= \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = 1 & \mu_{12} &= \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 1 = 2 \\ \mu_{21} &= \mu_0 + \alpha_2 + \beta_1 = 1 + 0 + 0 = 1 & \mu_{22} &= \mu_0 + \alpha_2 + \beta_2 = 1 + 0 + 1 = 2 \end{aligned}$$

Abbildung 11.4 B zeigt Visualisierungen dieses Erwartungswertmusters in der Balkendiagramm und Liniendiagrammform von Abbildung 11.3.

Beispiel (3)

Es seien $\mu_0 := 1, \alpha_2 := 1$ und $\beta_2 := 1$, die Effektparameter für die Haupteffekt beider Faktoren seien also von Null verschieden. Dann ergibt sich für die Gruppenerwartungswerte

$$\begin{aligned} \mu_{11} &= \mu_0 + \alpha_1 + \beta_1 = 1 + 0 + 0 = 1 & \mu_{12} &= \mu_0 + \alpha_1 + \beta_2 = 1 + 0 + 1 = 2 \\ \mu_{21} &= \mu_0 + \alpha_2 + \beta_1 = 1 + 1 + 0 = 2 & \mu_{22} &= \mu_0 + \alpha_2 + \beta_2 = 1 + 1 + 1 = 3 \end{aligned}$$

Abbildung 11.4 C zeigt Visualisierungen dieses Erwartungswertmusters in der Balkendiagramm und Liniendiagrammform von Abbildung 11.3.

In Hinblick auf die Liniendiagrammform der Gruppenerwartungswerte der additiven ZVA fällt auf, dass die den Levels von Faktor A entsprechenden Linien immer parallel sind.

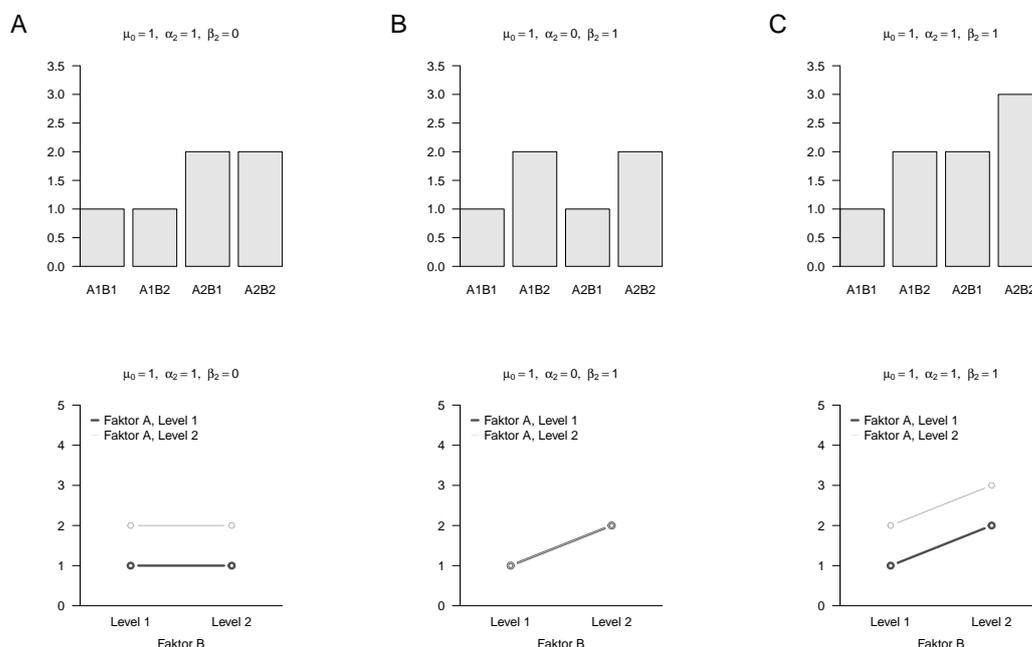


Abbildung 11.4 Gruppenerwartungswerte der Beispielparameterszenarien in Balken und Liniendiagrammform im Modell der additiven ZVA.

Für den Fall der 2×2 ZVA gibt Definition 11.2 abschließend das Modell der additiven ZVA mit Referenzgruppe inklusive seiner Designmatrixform an.

Definition 11.2 (Modell der additiven 2×2 ZVA mit Referenzgruppe). v_{ijk} mit $i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}$ 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 2×2 ZVA Anwendungsszenario modelliert. Dann hat mit $\sigma^2 > 0$ das *Modell der additiven 2×2 ZVA mit Referenzgruppe* die strukturelle Form

$$v_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}, \quad (11.5)$$

die Datenverteilungsform

$$v_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.v. für } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij} \quad (11.6)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j \text{ für } i = 1, 2, j = 1, 2 \text{ mit } \alpha_1 := \beta_1 := 0, \quad (11.7)$$

sowie die Designmatrixform

$$v \sim N(X\beta, \sigma^2 I_n), \quad (11.8)$$

wobei

$$v := \begin{pmatrix} v_{111} \\ \vdots \\ v_{11n_{11}} \\ v_{121} \\ \vdots \\ v_{12n_{12}} \\ v_{211} \\ \vdots \\ v_{21n_{21}} \\ v_{221} \\ \vdots \\ v_{22n_{22}} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times 3}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0. \quad (11.9)$$

•

11.2.2 Modell der ZVA mit Interaktion

In der ZVA mit Interaktion möchte man die Gruppenerwartungswerte μ_{ij} mit $i = 1, \dots, I$ für die Level von Faktor A und $j = 1, \dots, J$ für die Level von Faktor B als Summe eines gruppenunspezifischen Erwartungswertes, der Effekte der Level von Faktor A und Faktor B und der Interaktion der Level der Faktoren modellieren. Wir bezeichnen dazu den gruppenunspezifischen Erwartungswertparameter mit μ_0 , den Effekt von Level i von Faktor A mit α_i , den Effekt von Level j von Faktor B mit β_j , und die Interaktion von Level i von Faktor A mit Level j von Faktor B mit γ_{ij} . Dann ergibt sich zum Beispiel für $I := J := 2$

$$\begin{array}{l|l} \mu_{11} := \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} & \mu_{12} := \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} \\ \mu_{21} := \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} \end{array}$$

Wie in der die additiven ZVA ist diese Darstellung der Gruppenerwartungswerte μ_{ij} multipel überparameterisiert, speziell werden die vier Erwartungswertparameter $\mu_{11}, \mu_{12}, \mu_{21}$ und μ_{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 γ_{22} dargestellt. Um eine eindeutige Darstellung der μ_{ij} zu gewährleisten, bietet sich auch hier die Restriktion an, den Effekt des ersten Levels jedes Faktors und jeder Interaktion als Null zu definieren

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \dots, I, j = 1, \dots, J \quad (11.10)$$

und damit die Faktorlevelkombination A1B1 als Referenzgruppe zu etablieren. Es ergibt sich somit zum Beispiel für für $I := J := 2$

$$\begin{array}{c|c} \mu_{11} := \mu_0 & \mu_{12} := \mu_0 + \beta_2 \\ \hline \mu_{21} := \mu_0 + \alpha_2 & \mu_{22} := \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} \end{array}$$

Hier werden nun also die vier Erwartungswertparameter $\mu_{11}, \mu_{12}, \mu_{21}$ und μ_{22} durch die vier Effektparameter μ_0, α_2, β_2 und γ_{22} dargestellt. Natürlich ändern sich auch bei dieser Effektdarstellung des Modells der 2×2 ZVA mit Interaktion und Referenzgruppe die Interpretationen der Parameter $\mu_0, \alpha_2, \beta_2, \gamma_{22}$ im Vergleich zum überparameterisierten Fall ohne Referenzgruppe: μ_0 entspricht dem Erwartungswert der Faktorlevelkombination A1B1, α_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A, β_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B und γ_{22} der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B im Unterschied zum Übergang von Level 1 zu Level 2 von Faktor A. Wir fassen obige Überlegungen in folgender Definition, wobei wir wiederum auf die Angabe einer allgemeinen Designmatrixform verzichten wollen.

Definition 11.3 (Modell der ZVA mit Interaktion und Referenzgruppe). v_{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

$$v_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.11)$$

und die Datenverteilungsform

$$v_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. für } i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij} \quad (11.12)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j + \gamma_{ij} \quad (11.13)$$

sowie

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \dots, I, j = 1, \dots, J \quad (11.14)$$

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 11.5 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

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1 \quad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 0 + 0 = 1$$

$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 0 + 0 = 1 \quad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 0 + 0 + 2 = 3$$

Abbildung 11.5 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

$$\begin{aligned}\mu_{11} &= \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1 & \mu_{12} &= \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 1 + 0 = 2 \\ \mu_{21} &= \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 1 + 0 = 2 & \mu_{22} &= \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 1 + 1 - 2 = 1\end{aligned}$$

Abbildung 11.5 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

$$\begin{aligned}\mu_{11} &= \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1 & \mu_{12} &= \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 0 + 0 = 1 \\ \mu_{21} &= \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 1 + 0 + 0 = 2 & \mu_{22} &= \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 1 + 0 + 1 = 3\end{aligned}$$

Abbildung 11.5 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

$$\begin{aligned}\mu_{11} &= \mu_0 + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 0 + 0 + 0 = 1 & \mu_{12} &= \mu_0 + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 0 + 1 + 0 = 2 \\ \mu_{21} &= \mu_0 + \alpha_2 + \beta_1 + \gamma_{21} = 1 + 0 + 0 + 0 = 1 & \mu_{22} &= \mu_0 + \alpha_2 + \beta_2 + \gamma_{22} = 1 + 0 + 1 + 1 = 3\end{aligned}$$

Abbildung 11.5 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 immer nicht parallel verlaufen.

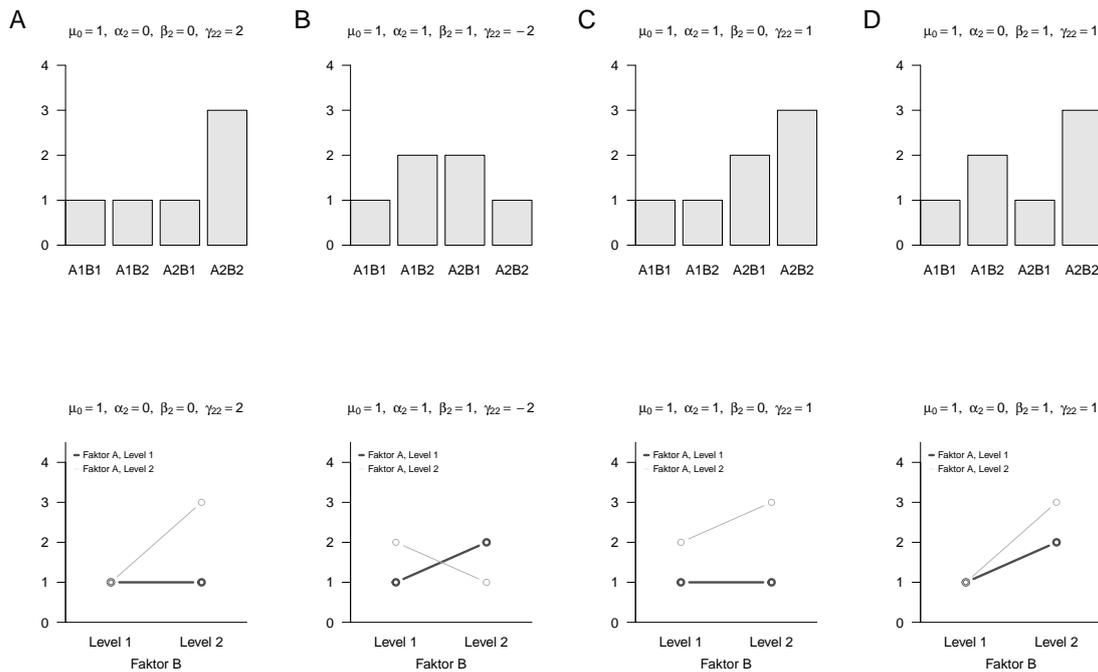


Abbildung 11.5 Gruppenerwartungswerte der Beispielparameterszenarien in Balken- und Liniendiagrammform im Modell der ZVA mit Interaktion.

Abschließend geben wir in Definition 11.4 den Spezialfall des Modells der 2×2 ZVA mit Interaktion und Referenzgruppe inklusiver seiner Designmatrixform an.

Definition 11.4 (Modell der 2×2 ZVA mit Interaktion und Referenzgruppe). v_{ijk} mit $i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}$ 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 2×2 ZVA Anwendungsszenario modelliert. Dann hat das Modell der 2×2 ZVA mit Interaktion und Referenzgruppe für $\sigma^2 > 0$ die strukturelle Form

$$v_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. f\"ur } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij}, \quad (11.15)$$

die Datenverteilungsform

$$v_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. f\"ur } i = 1, 2, j = 1, 2, k = 1, \dots, n_{ij} \quad (11.16)$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j + \gamma_{ij} \quad (11.17)$$

sowie

$$\alpha_1 := \beta_1 := \gamma_{11} := \gamma_{12} := 0 \text{ f\"ur } i = 1, \dots, I, j = 1, \dots, J \quad (11.18)$$

und f\"ur $n := \sum_{i=1}^2 \sum_{j=1}^2 n_{ij}$ die Designmatrixform

$$v \sim N(X\beta, \sigma^2 I_n), \text{ mit} \quad (11.19)$$

$$v := \begin{pmatrix} v_{111} \\ \vdots \\ v_{11n_{11}} \\ v_{121} \\ \vdots \\ v_{12n_{12}} \\ v_{211} \\ \vdots \\ v_{21n_{21}} \\ v_{221} \\ \vdots \\ v_{22n_{22}} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times 4}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (11.20)$$

•

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

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_{16}, \sigma^2 I_{16}) \quad (11.21)$$

mit

$$v := \begin{pmatrix} v_{111} \\ v_{112} \\ v_{113} \\ v_{114} \\ v_{121} \\ v_{122} \\ v_{123} \\ v_{124} \\ v_{211} \\ v_{212} \\ v_{213} \\ v_{214} \\ v_{221} \\ v_{222} \\ v_{223} \\ v_{224} \end{pmatrix}, X := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{16 \times 4}, \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0. \quad (11.22)$$

Folgender **R** Code erlaubt die Realisierung von Daten im Falle dieses Beispiels.

```

1 # Modellformulierung
2 library(MASS) # Multivariate Normalverteilung
3 I = 2 # Anzahl Level Faktor A
4 J = 2 # Anzahl Level Faktor B
5 n_ij = 4 # Anzahl von Datenpunkten der i,jten Gruppe
6 n = I*J*n_ij # Anzahl Datenpunkte
7 p = 1 + (I-1)+(J-1)+(I*J-3) # Anzahl Parameter
8 D = matrix(c(1,0,0,0, # Prototypische Designmatrix für balancierte Designs
9 1,0,1,0,
10 1,1,0,0,
11 1,1,1,1),
12 nrow = p,
13 byrow = TRUE)
14 C = matrix(rep(1,n_ij),nrow = n_ij) # Prototypischer Zellenvektor für balancierte Designs
15 X = kronecker(D,C) # Kroneckerprodukt Designmatrix Erzeugung für balancierte Designs

```

```

16 I_n = diag(n) # n x n Einheitsmatrix
17 beta = matrix(c(1,1,1,1), nrow = p) # \beta = (\mu_0, \alpha_2, \alpha_3, \alpha_4)
18 sigsqr = 10 # \sigma^2
19 y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
    
```

11.3 Modellschätzung

Wir betrachten im Folgenden die Betaparameterschätzer im additiven 2×2 ZVA Modell mit Referenzgruppe und im 2×2 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 2×2 ZVA Modell mit Referenzgruppe.

Theorem 11.1 (Betaparameterschätzung im additiven 2×2 ZVA Modell mit Referenzgruppe). *Gegeben sei die Designmatrixform eines balancierten additiven 2×2 ZVA Modells mit Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer*

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4}\bar{v}_{11} + \frac{1}{4}(\bar{v}_{12} + \bar{v}_{21}) - \frac{1}{4}\bar{v}_{22} \\ \frac{1}{2}(\bar{v}_{21} + \bar{v}_{22}) - \frac{1}{2}(\bar{v}_{11} + \bar{v}_{12}) \\ \frac{1}{2}(\bar{v}_{12} + \bar{v}_{22}) - \frac{1}{2}(\bar{v}_{11} + \bar{v}_{21}) \end{pmatrix}, \tag{11.23}$$

wobei

$$\bar{v}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} v_{ijk} \text{ für } 1 \leq i, j \leq 2 \tag{11.24}$$

das Stichprobenmittel der i, j ten Gruppe des 2×2 ZVA Designs bezeichnet.

Beweis. Wir bestimmen zunächst $X^T v$, $X^T X$ und $(X^T X)^{-1}$ bei konstantem n_{ij} für $1 \leq i, j \leq 2$.

$$X^T v = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} v_{111} \\ \vdots \\ v_{11n_{11}} \\ v_{121} \\ \vdots \\ v_{12n_{12}} \\ v_{211} \\ \vdots \\ v_{21n_{21}} \\ v_{221} \\ \vdots \\ v_{22n_{22}} \end{pmatrix} \tag{11.25}$$

$$= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} \end{pmatrix}.$$

$$\begin{aligned}
X^T X &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix} \\
&= n_{ij} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}
\end{aligned} \tag{11.26}$$

Ohne Beweis halten wir weiterhin fest, dass

$$(X^T X)^{-1} = n_{ij} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}^{-1} = \frac{1}{n_{ij}} \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \tag{11.27}$$

Es ergibt sich also

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \frac{1}{n_{ij}} \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} \end{pmatrix}. \tag{11.28}$$

Damit ergibt sich dann

$$\begin{aligned}
\hat{\mu}_0 &= \frac{1}{n_{ij}} \left(\frac{3}{4} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} - \frac{1}{2} \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} - \frac{1}{2} \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} \right) \\
&= \frac{1}{n_{ij}} \left(\frac{3}{4} \sum_{k=1}^{n_{11}} v_{11k} + \frac{3}{4} \sum_{k=1}^{n_{12}} v_{12k} + \frac{3}{4} \sum_{k=1}^{n_{21}} v_{21k} + \frac{3}{4} \sum_{k=1}^{n_{22}} v_{22k} \right) \\
&\quad + \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{k=1}^{n_{21}} v_{21k} - \frac{1}{2} \sum_{k=1}^{n_{22}} v_{22k} - \frac{1}{2} \sum_{k=1}^{n_{12}} v_{12k} - \frac{1}{2} \sum_{k=1}^{n_{11}} v_{11k} \right) \\
&= \frac{1}{n_{ij}} \left(\frac{3}{4} \sum_{k=1}^{n_{11}} v_{11k} + \frac{1}{4} \sum_{k=1}^{n_{12}} v_{12k} + \frac{1}{4} \sum_{k=1}^{n_{21}} v_{21k} - \frac{1}{4} \sum_{k=1}^{n_{22}} v_{22k} \right) \\
&= \frac{3}{4} \bar{v}_{11} + \frac{1}{4} (\bar{v}_{12} + \bar{v}_{21}) - \frac{1}{4} \bar{v}_{22}
\end{aligned} \tag{11.29}$$

sowie

$$\begin{aligned}
\hat{\alpha}_2 &= \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} + \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} \right) \\
&= \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{k=1}^{n_{11}} v_{11k} - \frac{1}{2} \sum_{k=1}^{n_{12}} v_{12k} - \frac{1}{2} \sum_{k=1}^{n_{21}} v_{21k} - \frac{1}{2} \sum_{k=1}^{n_{22}} v_{22k} + \sum_{k=1}^{n_{21}} v_{21k} + \sum_{k=1}^{n_{22}} v_{22k} \right) \\
&= \frac{1}{n_{ij}} \left(-\frac{1}{2} \sum_{k=1}^{n_{11}} v_{11k} - \frac{1}{2} \sum_{k=1}^{n_{12}} v_{12k} + \frac{1}{2} \sum_{k=1}^{n_{21}} v_{21k} + \frac{1}{2} \sum_{k=1}^{n_{22}} v_{22k} \right) \\
&= \frac{1}{2} (\bar{v}_{21} + \bar{v}_{22}) - \frac{1}{2} (\bar{v}_{11} + \bar{v}_{12})
\end{aligned} \tag{11.30}$$

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.

```

1 # Datenreformatierung
2 fname = file.path("../_data/411-Zweifaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE) # Datensatz
4 A1B1 = D$dBDI[D$Setting == "F2F" & D$Variant == "MND"] # Face-to-face, mindfulness
5 A1B2 = D$dBDI[D$Setting == "F2F" & D$Variant == "EXC"] # Face-to-face, exercise
6 A2B1 = D$dBDI[D$Setting == "ONL" & D$Variant == "MND"] # Online, mindfulness
7 A2B2 = D$dBDI[D$Setting == "ONL" & D$Variant == "EXC"] # Online, exercise
8
9 # Datenmatrix für Gruppenmittelwerte
10 n_ij = length(A1B1) # Anzahl Datenpunkte pro Gruppe
11 Y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n_ij) # Datenmatrix
12 bar_y = colMeans(Y) # Zellenmittelwerte
13
14 # Modellschätzung
15 I = 2 # Anzahl Level Faktor A (Therapiesetting)
16 J = 2 # Anzahl Level Faktor B (Therapievariante)
17 n = I*J*n_ij # Anzahl Datenpunkte
18 p = 1 + (I-1)+(J-1)+(I*J-3) # Anzahl Parameter
19 D = matrix(c(1,0,0,
20             1,0,1,
21             1,1,0,
22             1,1,1), nrow = p, byrow = TRUE) # Prototypische Designmatrix für balancierte Designs
23 C = matrix(rep(1,n_ij),nrow = n_ij) # Prototypischer Zellenvektor für balancierte Designs
24 X = kronecker(D,C) # Kroneckerprodukt Designmatrix
25 y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
26 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
27 eps_hat = y - X %*% beta_hat # Residuenvektor
28 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
29 cat("hat{beta} : ", beta_hat, # Ausgabe
30     "\nhat{sigsqr} : ", sigsq_hat,
31     "\ny_11,y_12,y_21,y_22 : ", bar_y,
32     "\n3/4y_11 + 1/4(y_12 + y_21) - 1/4y_22 : ", (3/4)*bar_y[1]+(1/4)*(bar_y[2]+bar_y[3])-(1/4)*bar_y[4],
33     "\n1/2(y_21 + y_22) - 1/2(y_11 + y_12) : ", (1/2)*(bar_y[3]+bar_y[4])-(1/2)*(bar_y[1] + bar_y[2]),
34     "\n1/2(y_12 + y_22) - 1/2(y_11 + y_21) : ", (1/2)*(bar_y[2]+bar_y[4])-(1/2)*(bar_y[1] + bar_y[3]))

```

```

hat{beta} : 10.14583 0.2916667 4.041667
hat{sigsqr} : 6.074337
y_11,y_12,y_21,y_22 : 10.5 13.83333 10.08333 14.83333
3/4y_11 + 1/4(y_12 + y_21) - 1/4y_22 : 10.14583
1/2(y_21 + y_22) - 1/2(y_11 + y_12) : 0.2916667
1/2(y_12 + y_22) - 1/2(y_11 + y_21) : 4.041667

```

Für die Betaparameterschätzung im 2×2 ZVA Modell mit Interaktion und Referenzgruppe ergibt sich folgendes Theorem.

Theorem 11.2 (Betaparameterschätzung im 2×2 ZVA Modell mit Interaktion und Referenzgruppe). *Gegeben sei die Designmatrixform eines balancierten 2×2 ZVA Modells mit Interaktion und Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer*

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \begin{pmatrix} \bar{v}_{11} \\ \bar{v}_{21} - \bar{v}_{11} \\ \bar{v}_{12} - \bar{v}_{11} \\ \bar{v}_{11} + \bar{v}_{22} - \bar{v}_{12} - \bar{v}_{21} \end{pmatrix}, \quad (11.31)$$

wobei

$$\bar{v}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} v_{ijk} \text{ für } 1 \leq i, j \leq 2 \quad (11.32)$$

das Stichprobenmittel der i, j ten Gruppe des 2×2 ZVA Designs bezeichnet.

Beweis. Wir bestimmen zunächst $X^T v$, $X^T X$ und $(X^T X)^{-1}$ bei konstantem n_{ij} für $1 \leq i, j \leq 2$.

$$\begin{aligned}
 X^T v &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} v_{111} \\ \vdots \\ v_{11n_{11}} \\ v_{121} \\ \vdots \\ v_{12n_{12}} \\ v_{211} \\ \vdots \\ v_{21n_{21}} \\ v_{221} \\ \vdots \\ v_{22n_{22}} \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} \\ \sum_{k=1}^{n_{22}} v_{22k} \end{pmatrix}.
 \end{aligned} \tag{11.33}$$

$$\begin{aligned}
 X^T X &= \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \\
 &= n_{ij} \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.
 \end{aligned} \tag{11.34}$$

Ohne Beweis halten wir weiterhin fest, dass

$$(X^T X)^{-1} = n_{ij} \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}^{-1} = \frac{1}{n_{ij}} \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 4 \end{pmatrix}. \tag{11.35}$$

Es ergibt sich also

$$\hat{\beta} = \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \frac{1}{n_{ij}} \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 2 & 1 & -2 \\ -1 & 1 & 2 & -2 \\ 1 & -2 & -2 & 4 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} \\ \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} \\ \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} \\ \sum_{k=1}^{n_{22}} v_{22k} \end{pmatrix}. \tag{11.36}$$

Damit ergibt sich dann

$$\begin{aligned}
 \hat{\mu}_0 &= \frac{1}{n_{ij}} \left(\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} - \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} - \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} + \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{11}} v_{11k} + \sum_{k=1}^{n_{12}} v_{12k} + \sum_{k=1}^{n_{21}} v_{21k} + \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &+ \frac{1}{n_{ij}} \left(- \sum_{k=1}^{n_{21}} v_{21k} - \sum_{k=1}^{n_{22}} v_{22k} - \sum_{k=1}^{n_{12}} v_{12k} - \sum_{k=1}^{n_{22}} v_{22k} + \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &= \frac{1}{n_{11}} \sum_{k=1}^{n_{11}} v_{11k} \\
 &= \bar{v}_{11}
 \end{aligned} \tag{11.37}$$

sowie

$$\begin{aligned}
 \hat{\alpha}_2 &= \frac{1}{n_{ij}} \left(- \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} + 2 \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} + 1 \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} - 2 \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left(- \sum_{k=1}^{n_{11}} v_{11k} - \sum_{k=1}^{n_{12}} v_{12k} - \sum_{k=1}^{n_{21}} v_{21k} - \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &+ \frac{1}{n_{ij}} \left(2 \sum_{k=1}^{n_{21}} v_{21k} + 2 \sum_{k=1}^{n_{22}} v_{22k} + \sum_{k=1}^{n_{12}} v_{12k} + \sum_{k=1}^{n_{22}} v_{22k} - 2 \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{21}} v_{21k} - \sum_{k=1}^{n_{11}} v_{11k} \right) \\
 &= \bar{v}_{21} - \bar{v}_{11}
 \end{aligned} \tag{11.38}$$

und analog für $\hat{\beta}_2$. Schließlich ergibt sich

$$\begin{aligned}
 \hat{\gamma}_{22} &= \frac{1}{n_{ij}} \left(\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} v_{ijk} - 2 \sum_{j=1}^2 \sum_{k=1}^{n_{2j}} v_{2jk} - 2 \sum_{i=1}^2 \sum_{k=1}^{n_{i2}} v_{i2k} + 4 \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{11}} v_{11k} + \sum_{k=1}^{n_{12}} v_{12k} + \sum_{k=1}^{n_{21}} v_{21k} + \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &+ \frac{1}{n_{ij}} \left(-2 \sum_{k=1}^{n_{21}} v_{21k} - 2 \sum_{k=1}^{n_{22}} v_{22k} - 2 \sum_{k=1}^{n_{12}} v_{12k} - 2 \sum_{k=1}^{n_{22}} v_{22k} + 4 \sum_{k=1}^{n_{22}} v_{22k} \right) \\
 &= \frac{1}{n_{ij}} \left(\sum_{k=1}^{n_{11}} v_{11k} + \sum_{k=1}^{n_{22}} v_{22k} - \sum_{k=1}^{n_{12}} v_{12k} - \sum_{k=1}^{n_{21}} v_{21k} \right) \\
 &= \bar{v}_{11} + \bar{v}_{22} - \bar{v}_{12} - \bar{v}_{21}.
 \end{aligned} \tag{11.39}$$

□

Auch hier verdeutlichen wir den in obigem Theorem formulierten Zusammenhang zwischen den Betaparameterschätzerkomponenten und den gruppenspezifischen Stichprobenmitteln anhand des Beispieldatensatzes mithilfe folgenden **R** Codes.

```

1 # Datenreformatierung
2 fname = file.path("../data/411-Zweifaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE) # Datensatz
4 A1B1 = D$dBDI[D$Setting == "F2F" & D$Variant == "MND"] # Face-to-face, mindfulness
5 A1B2 = D$dBDI[D$Setting == "F2F" & D$Variant == "EXC"] # Face-to-face, exercise
6 A2B1 = D$dBDI[D$Setting == "ONL" & D$Variant == "MND"] # Online, mindfulness
7 A2B2 = D$dBDI[D$Setting == "ONL" & D$Variant == "EXC"] # Online, exercise
8
9 # Datenmatrix für Gruppenmittelwerte
10 n_ij = length(A1B1) # Anzahl Datenpunkte pro Gruppe
11 Y = matrix(c(A1B1, A1B2, A2B1, A2B2), nrow = n_ij) # Datenmatrix
12 bar_y = colMeans(Y) # Zellenmittelwerte
13
14 # Modellschätzung

```

```

15 I      = 2                                # Anzahl Level Faktor A (Therapiesetting)
16 J      = 2                                # Anzahl Level Faktor B (Therapievariante)
17
18 # Datenmatrix für Gruppenmittelwerte
19 n_ij    = length(A1B1)                    # Anzahl Datenpunkte pro Gruppe
20 Y       = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n_ij) # Datenmatrix
21 bar_y   = colMeans(Y)                    # Zellenmittelwerte
22
23 # Modellschätzung
24 I       = 2                                # Anzahl Level Faktor A (Therapie)
25 J       = 2                                # Anzahl Level Faktor B (Alter)
26 n       = I*J*n_ij                        # Anzahl Datenpunkte
27 p       = 1 + (I-1)+(J-1)+(I*J-3)        # Anzahl Parameter
28 D       = matrix(c(1,0,0,0,              # Prototypische Designmatrix für balancierte Designs
29                 1,0,1,0,
30                 1,1,0,0,
31                 1,1,1,1), nrow = p, byrow = TRUE)
32 C       = matrix(rep(1,n_ij),nrow = n_ij) # Prototypischer Zellenvektor für balancierte Designs
33 X       = kronecker(D,C)                  # Kroneckerprodukt Designmatrix
34 y       = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
35 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
36 eps_hat  = y - X %*% beta_hat            # Residuenvektor
37 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
38 cat("hat{beta}          : ", beta_hat,
39     "\nhat{sigsqr}       : ", sigsq_hat,
40     "\ny_11,y_12,y_21,y_22 : ", bar_y,
41     "\ny_11              : ", bar_y[1],
42     "\ny_21 - y_11       : ", bar_y[3]-bar_y[1],
43     "\ny_12 - y_11       : ", bar_y[2]-bar_y[1],
44     "\ny_11 + y_22 - y_12 + y_21 : ", bar_y[1]+bar_y[4]-bar_y[3]-bar_y[2])

hat{beta}          : 10.5 -0.4166667 3.333333 1.416667
hat{sigsqr}       : 5.9375
y_11,y_12,y_21,y_22 : 10.5 13.833333 10.083333 14.833333
y_11              : 10.5
y_21 - y_11       : -0.4166667
y_12 - y_11       : 3.333333
y_11 + y_22 - y_12 + y_21 : 1.416667

```

11.4 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 2×2 ZVA mit Referenzgruppe und (2) die Evaluation der Interaktion im Modell der 2×2 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.

11.4.1 Evaluation von Haupteffekten im additiven Modell der 2×2 ZVA

Theorem 11.3 (Teststatistiken für Haupteffekte).

Es sei

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{11.40}$$

die Designmatrixform des additiven Modells der 2×2 ZVA mit Referenzgruppe, wobei die $n \times 1$ Spalten von X bezeichnet seien durch

$$X := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2}) \in \mathbb{R}^{n \times 3} \tag{11.41}$$

Dann gelten

(A) Eine F-Teststatistik für den Haupteffekt von Faktor A ist die F-Statistik des ALMs mit

$$X_A := (X_{\mu_0} \quad X_{\beta_2} \quad X_{\alpha_2}) \in \mathbb{R}^{n \times 3}, \beta_A := \begin{pmatrix} \mu_0 \\ \beta_2 \\ \alpha_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } p_0 := 2, p_1 := 1. \quad (11.42)$$

(B) Eine F-Teststatistik für den Haupteffekt von Faktor B ist die F-Statistik des ALMs mit

$$X_B := (X_{\mu_0} \quad X_{\alpha_2} \quad X_{\beta_2}) \in \mathbb{R}^{n \times 3}, \beta_B := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } p_0 := 2, p_1 := 1. \quad (11.43)$$

Definition 11.5 (Testhypothesen und Tests für Haupteffekte).

Gegeben sei das Modell der additiven 2×2 ZVA und die F-Teststatistiken für die Haupteffekte von Faktor A und B seien mit F_A und F_B bezeichnet und wie oben definiert. Dann gelten folgende Definitionen:

(A) Der kritische Wert-basierte Test

$$\phi_A(v) := 1_{\{F_A \geq k\}} \text{ mit Nullhypothese } H_0^A : \alpha_2 = 0 \quad (11.44)$$

definiert den *F-Test des Haupteffekts von Faktor A* im Modell mit X_A, β_A .

(B) Der kritische Wert-basierte Test

$$\phi_B(v) := 1_{\{F_B \geq k\}} \text{ mit Nullhypothese } H_0^B : \beta_2 = 0 \quad (11.45)$$

definiert den *F-Test des Haupteffekts von Faktor B* im Modell mit X_B, β_B .

•

Theorem 11.4 (Testumfangkontrolle und p-Werte für Haupteffekte).

Mit obigen Definitionen und der KVF φ der f -Verteilung gelten:

(A) ϕ_A ist ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0}^A := \varphi^{-1}(1 - \alpha_0; 1, n - 3). \quad (11.46)$$

Der zu einem beobachteten Wert f_A von F_A assoziierte p -Wert ist gegeben durch

$$p_A\text{-Wert} := 1 - \varphi(f_A; 1, n - 3). \quad (11.47)$$

(B) ϕ_B ist ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0}^B := \varphi^{-1}(1 - \alpha_0; 1, n - 3). \quad (11.48)$$

Der zu einem beobachteten Wert f_B von F_B assoziierte p -Wert ist gegeben durch

$$p_B\text{-Wert} := 1 - \varphi(f_B; 1, n - 3). \quad (11.49)$$

Anwendungsbeispiel

Die Anwendung des oben formulierten Evaluationsverfahrens für die Haupteffekte im Modell der additiven 2×2 ZVA mit Referenzgruppe anhand des Beispieldatensatzes demonstriert folgender **R** Code. Für den Haupteffekt des Faktors **Settings** ergibt sich dabei ein Wert der F -Statistik von $F_A = 0.17$ bei einem kritischen Wert von $k_{0.05} = 4.05$. Die Nullhypothese eines wahren, aber unbekanntem, Effektparameterwertes von Null für den Faktor **Setting** würde im Lichte der in Abbildung 11.3 visualisierten Daten also nicht verworfen werden. Für den Haupteffekt des Faktors **Variant** dagegen ergibt ein Wert der F -Statistik von $F_B = 33.0$ bei dem gleichen kritischen Wert von $k_{0.05} = 4.05$. Die Nullhypothese eines wahren, aber unbekanntem, Effektparameterwertes von Null für den Faktor **Variant** würde im Lichte der in Abbildung 11.3 visualisierten Daten mit einem Signifikanzlevel von $\alpha_0 := 0.05$ also verworfen werden.

```

1 # Modellevaluation
2 I = 2 # Anzahl Level Faktor A (Therapie)
3 J = 2 # Anzahl Level Faktor B (Alter)
4 n_ij = length(A1B1) # balanciertes ANOVA Design
5 n = I*J*n_ij # Anzahl Datenpunkte
6 p = 3 # Anzahl Parameter vollständiges Modell
7 y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
8 D = matrix(c(1,0,0,1,0,1,1,1,1,0,1,1,1), # Prototypische Designmatrix
9 nrow = I*J,byrow=TRUE)
10 C = matrix(rep(1,n_ij),nrow = n_ij) # Prototypischer Zellenvektor für balancierte Designs
11 X = kronecker(D,C) # ZVA Kroneckerprodukt Designmatrix
12 XH = list(X[,c(1,3,2)], X) # Modellvarianten
13 alpha_0 = 0.05 # Signifikanzlevel
14 Eff = rep(NA,n,2) # F-Teststatistik Arrayinitialisierung
15 k_alpha_0 = rep(NA,n,2) # Kritischer Wert Arrayinitialisierung
16 phi = rep(NA,n,2) # Testwert Arrayinitialisierung
17 p_vals = rep(NA,n,2) # p-Wert Arrayinitialisierung
18 for(i in 1:2){ # Iteration über Modellvarianten
19 X = XH[[i]] # Designmatrix vollständiges Modell
20 X_0 = X[,-3] # Designmatrix reduziertes Modell
21 p = ncol(X) # Anzahl Parameter vollständiges Modell
22 p_0 = ncol(X_0) # Anzahl Parameter reduziertes Modell
23 p_1 = p - p_0 # Anzahl zusätzlicher Parameter im vollst. Modell
24 beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y # Betaparameterschätzer reduziertes Modell
25 beta_hat = solve(t(X) %*%X )%*%t(X) %*%y # Betaparameterschätzer vollständiges Modell
26 eps_hat_0 = y - X_0%*%beta_hat_0 # Residuenvektor reduziertes Modell
27 eps_hat = y - X%*%beta_hat # Residuenvektor vollständiges Modell
28 eh0_eh0 = t(eps_hat_0) %*% eps_hat_0 # RQS reduziertes Modell
29 eh_eh = t(eps_hat) %*% eps_hat # RQS vollständiges Modell
30 sigsq_hat = eh_eh/(n-p) # Varianzparameterschätzer vollst. Modell
31 Eff[i] = ((eh0_eh0-eh_eh)/p_1)/sigsq_hat # F-Statistik
32 k_alpha_0[i] = qf(1-alpha_0, p_1, n-p) # kritischer Wert
33 if(Eff[i] >= k_alpha_0[i]){ phi[i] = 1 } # H_0 Ablehnen
34 else { phi[i] = 0 } # H_A Annehmen
35 p_vals[i] = 1 - pf(Eff[i], p_1,n-p) # p-Wert
36 }
37 data.frame("f"= Eff,"k" = k_alpha_0,"phi"= phi, "p-Wert" = p_vals, row.names = c("Setting", "Variant"))

      f      k phi      p.Wert
Setting 0.1718762 4.056612 0 6.804195e-01
Variant 33.0037415 4.056612 1 7.456844e-07

```

Folgender **R** Code demonstriert die gleiche Analyse mithilfe der **R** Funktion `aov()`.

```

1 # Dateneinlesen
2 fname = file.path("../data/411-Zweifaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # R's aov Funktion
6 res.aov = aov(dBDI ~ Setting + Variant, data = D) # Modellformulierung und Modellschätzung
7 summary(res.aov) # Modellevaluation

      Df Sum Sq Mean Sq F value Pr(>F)
Setting  1  1.02  1.02  0.172  0.68
Variant  1 196.02 196.02 33.004 7.46e-07 ***
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()`.

```

1 # Dateneinlesen
2 fname = file.path("../data/411-Zweifaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # R's lm (linear model) und anova() Funktionen
6 glm = lm(dBDI ~ Setting + Variant, data = D) # Modellformulierung und Modellschätzung
7 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	196.021	196.021	33.0037	7.457e-07 ***
Residuals	45	267.271	5.939		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

11.4.2 Evaluation der Interaktion im Modell der 2×2 ZVA mit Interaktion

Theorem 11.5 (Teststatistik für die Interaktion). *Es sei*

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{11.50}$$

die Designmatrixform des 2×2 ZVA Modells mit Interaktion und Referenzgruppe. Dann ist die F -Statistik mit $p_0 := 3$ und $p_1 := 1$ eine F -Teststatistik für die Interaktion von Faktor A und Faktor B .

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

$$\phi_{A \times B}(v) := 1_{\{F_{A \times B} \geq k\}} \text{ mit Nullhypothese } H_0^{A \times B} : \gamma_{22} = 0 \tag{11.51}$$

den F -Test der Interaktion von Faktor A und Faktor B .

•

Theorem 11.6 (Testumfangkontrolle und p-Wert für die Interaktion). *Mit obigen Definition und der KVF φ der f -Verteilung gilt, dass $\phi_{A \times B}$ ein Level- α_0 -Test mit Testumfang α_0 ist, wenn der kritische Wert definiert ist durch*

$$k_{\alpha_0}^{A \times B} := \varphi^{-1}(1 - \alpha_0; 1, n - 4). \tag{11.52}$$

Der zu einem beobachteten Wert $f_{A \times B}$ von $F_{A \times B}$ assoziierte p -Wert ist gegeben durch

$$p_{A \times B}\text{-Wert} := 1 - \varphi(f_{A \times B}; 1, n - 4). \tag{11.53}$$

Anwendungsbeispiel

Die Anwendung des oben formulierten Evaluationsverfahrens für die Interaktion der Faktoren im Modell der 2×2 ZVA mit Interaktion und Referenzgruppe anhand des Beispieldatensatzes demonstriert folgender **R** Code. Es ergibt sich dabei eine 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 unbekanntes, Effektparameterwertes von Null für den Interaktionseffekt würde im Lichte der in Abbildung 11.3 visualisierten Daten also nicht verworfen werden.

```

1 # Modellevaluation
2 I = 2 # Anzahl Level Faktor A (Therapie)
3 J = 2 # Anzahl Level Faktor B (Alter)
4 n_ij = length(A1B1) # balanciertes ANOVA Design
5 n = I*J*n_ij # Anzahl Datenpunkte
6 p = 4 # Anzahl Parameter vollständiges Modell
7 y = matrix(c(A1B1,A1B2,A2B1,A2B2), nrow = n) # Datenvektor
8 D = matrix(c(1,0,0,0,1,0,1,0, 1,1,0,0,1,1,1,1), # Prototypische Designmatrix
9 nrow = I*J,byrow=TRUE)
10 C = matrix(rep(1,n_ij),nrow = n_ij) # Prototypischer Zellenvektor für balancierte Designs
11 X = kronecker(D,C) # ZVA Kroneckerprodukt Designmatrix
12 XH = list(X[,c(1,3,2)], X) # Modellvarianten
13 alpha_0 = 0.05 # Signifikanzlevel
14 X = XH[[1]] # Designmatrix vollständiges Modell
15 X_0 = X[,-4] # Designmatrix reduziertes Modell
16 p = ncol(X) # Anzahl Parameter vollständiges Modell
17 p_0 = ncol(X_0) # Anzahl Parameter reduziertes Modell
18 p_1 = p - p_0 # Anzahl zusätzlicher Parameter im vollst. Modell
19 beta_hat_0 = solve(t(X_0)%*%X_0)%*%t(X_0)%*%y # Betaparameterschätzer reduziertes Modell
20 beta_hat = solve(t(X) %*%X )%*%t(X) %*%y # Betaparameterschätzer vollständiges Modell
21 eps_hat_0 = y-X_0%*%beta_hat_0 # Residuenvektor reduziertes Modell
22 eps_hat = y - X%*%beta_hat # Residuenvektor vollständiges Modell
23 eh_0_eh0 = t(eps_hat_0) %*% eps_hat_0 # RQS reduziertes Modell
24 eh_eh = t(eps_hat) %*% eps_hat # RQS vollständiges Modell
25 sigsq_hat = eh_eh/(n-p) # Varianzparameterschätzer vollst. Modell
26 f = ((eh_0_eh0-eh_eh)/p_1)/sigsqr_hat # F-Statistik
27 k_alpha_0 = qf(1-alpha_0, p_1, n-p) # kritischer Wert
28 if(f >= k_alpha_0){phi = 1} else {phi = 0} # Test
29 p_val = 1 - pf(f, p_1,n-p) # p-Wert
30 data.frame("f" = f,"k" = k_alpha_0," phi" = phi, "p-Wert" = p_val, row.names = c("Setting x Variant"))

```

```

          f          k X.phi    p.Wert
Setting x Variant 1.014035 4.061706      0 0.319444

```

Folgender R Code demonstriert die gleiche Analyse mithilfe der R Funktion aov().

```

1 # Dateneinlesen
2 fname = file.path("../_data/411-Zweifaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # R's aov Funktion
6 res.aov = aov(dBDI ~ Setting + Variant + Setting:Variant, data = D) # Modellformulierung und Modellschätzung
7 summary(res.aov) # Modellevaluation

```

```

          Df Sum Sq Mean Sq F value    Pr(>F)
Setting   1   1.02    1.02    0.172    0.680
Variant  1 196.02  196.02  33.014 7.97e-07 ***
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().

```

1 # Dateneinlesen
2 fname = file.path("../_data/411-Zweifaktorielle-Varianzanalyse.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # R's lm (linear model) und anova() Funktionen
6 glm = lm(dBDI ~ Setting + Variant + Setting:Variant, data = D) # Modellformulierung und Modellschätzung
7 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  1 196.021  196.021  33.0140 7.967e-07 ***
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

```

11.5 Literaturhinweise

Die Popularität varianzanalytischer Verfahren wird im Allgemeinen auf Fisher (1925b) und Fisher (1935) zurückgeführt. Everitt und Howell (2005) und Stigler (1986) geben einen kurzen und einen ausführlichen historischen Überblick, respektive.

11.6 Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario der zweifaktoriellen Varianzanalyse.
2. Aus wie vielen Datenpunkten besteht ein Datensatz eines 3×4 ZVA Designs mit 10 Datenpunkten pro Zelle?
3. Erläutern Sie die intuitive Bedeutung eines Haupteffektes in einem ZVA Design.
4. Erläutern Sie die intuitive Bedeutung einer Interaktion in einem ZVA Design.
5. Geben Sie die Definition des additiven Modells der ZVA mit Referenzgruppe wieder.
6. Erläutern Sie die Bedeutung der Parameter μ_0 , α_2 und β_2 im additiven Modell der ZVA mit Referenzgruppe.
7. Bestimmen Sie μ_{ij} für $\mu_0 := 2$, $\alpha_2 = -1$ und $\beta_2 := 3$ im additiven Modell der ZVA mit Referenzgruppe.
8. Geben Sie die Designmatrixform des Modells einer additiven 2×2 ZVA mit Referenzgruppe für $n_{ij} := 1$ an.
9. Geben Sie die Designmatrixform des Modells einer additiven 2×2 ZVA mit Referenzgruppe 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 μ_0 , α_2 , β_2 und γ_{22} im Modell der ZVA mit Interaktion und Referenzgruppe.
12. Geben Sie die Designmatrixform des Modells einer 2×2 ZVA mit Interaktion und Referenzgruppe für $n_{ij} := 1$ an.
13. Geben Sie die Designmatrixform des Modells einer 2×2 ZVA mit Interaktion und Referenzgruppe für $n_{ij} := 3$ an.
14. Geben Sie das Theorem zur Betaparameterschätzung im additiven 2×2 ZVA Modell mit Referenzgruppe wieder.
15. Geben Sie das Theorem zur Betaparameterschätzung im 2×2 ZVA Modell mit Interaktion und Referenzgruppe 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 sieht man eine Tendenz dafür, 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.

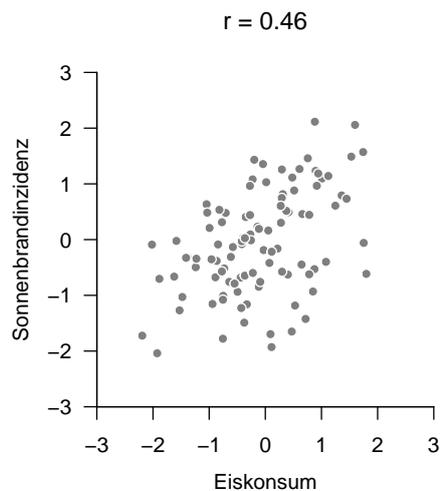


Abbildung 12.1 Beispielszenario zur Evaluation bedingter und partieller Korrelationen.

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*, wie zum Beispiel 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).

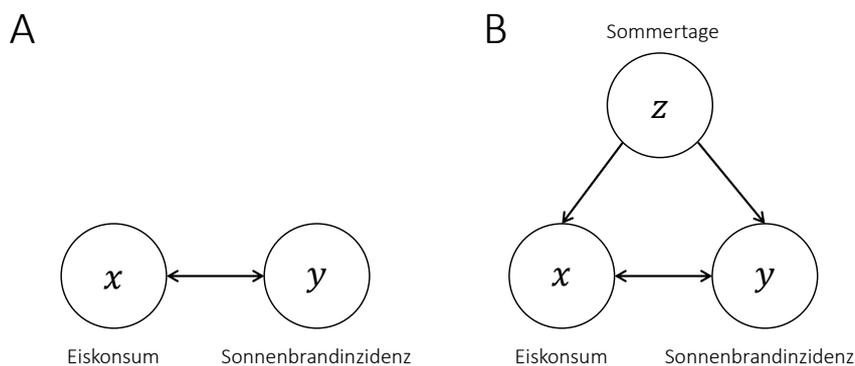


Abbildung 12.2 Zufallsvariablen im Beispielszenario

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 einer *bedingten Korrelation* 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 Kapitel 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 Kapitel 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. Aus Gründen der notationellen Einfachheit bezeichnen wir die in diesem Abschnitt betrachteten Zufallsvariablen mit x, y und z , nutzen im Gegensatz zur üblichen Gewohnheit für sie also lateinische, nicht griechische Buchstaben.

12.2 Bedingte Korelation

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

$$\rho(x, y|z) := \frac{\mathbb{C}(x, y|z)}{\mathbb{S}(x|z)\mathbb{S}(y|z)} \quad (12.1)$$

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. für $v := (x, y, z)^T$ gelte, dass

$$v \sim N(\mu, \Sigma) \quad (12.2)$$

mit

$$\mu := \begin{pmatrix} \mu_y \\ \mu_x \\ \mu_z \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 & \sigma_{x,z}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 & \sigma_{y,z}^2 \\ \sigma_{z,x}^2 & \sigma_{z,y}^2 & \sigma_z^2 \end{pmatrix} \quad (12.3)$$

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 4.8 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

$$\Sigma_{x,y} := \begin{pmatrix} \sigma_x^2 & \sigma_{x,y}^2 \\ \sigma_{y,x}^2 & \sigma_y^2 \end{pmatrix}, \Sigma_z := (\sigma_z^2) \text{ und } \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^T := \begin{pmatrix} \sigma_{x,z}^2 \\ \sigma_{y,z}^2 \end{pmatrix}, \quad (12.4)$$

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 4.8 ergibt sich der Kovarianzmatrixparameter des Zufallsvektors $(x, y)^T$ dann zu

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z} \Sigma_z^{-1} \Sigma_{z,(x,y)}. \quad (12.6)$$

Mit den Eigenschaften von multivariaten Normalverteilungen gilt dann, dass die Diagonaleinträge von $\Sigma_{x,y|z}$ den bedingten Varianzen von x und y gegeben z entsprechen und dass der Nichtdiagonaleintrag von $\Sigma_{x,y|z}$ 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 $(x, y, z)^T$ 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.

```

1 # Bedingte Korrelation bei Normalverteilung
2 S = matrix(c(.5,.9,          # \Sigma
3             .5, 1,.5,
4             .9,.5, 1), nrow = 3, byrow = TRUE)
5 rho_xy = S[1,2]/(sqrt(S[1,1])*sqrt(S[2,2])) # \rho(x,y)
6 S_xy_z = S[1:2,1:2] - S[1:2,3] %*% solve(S[2,2]) %*% S[3,1:2] # \Sigma_{x,y|z}
7 rho_xy_z = S_xy_z[1,2]/(sqrt(S_xy_z[1,1])*sqrt(S_xy_z[2,2])) # \rho(x,y|z)
8
9 # Ausgabe
10 cat("rho(x,y) : ", rho_xy,
11      "\nrho(x,y|z) : ", rho_xy_z)

```

```

rho(x,y) : 0.5
rho(x,y|z) : 0.1324532

```

Bedingte Korrelation und Korrelationen 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 multivariate 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

$$\Sigma_{x,y} := \begin{pmatrix} 1 & \rho(x, y) \\ \rho(y, x) & 1 \end{pmatrix}, \Sigma_z := (1) \text{ und } \Sigma_{(x,y),z} := \Sigma_{z,(x,y)}^T := \begin{pmatrix} \rho(x, z) \\ \rho(y, z) \end{pmatrix}, \quad (12.13)$$

so dass

$$\Sigma = \begin{pmatrix} \Sigma_{x,y} & \Sigma_{(x,y),z} \\ \Sigma_{z,(x,y)} & \Sigma_z \end{pmatrix}. \quad (12.14)$$

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

$$\Sigma_{x,y|z} = \Sigma_{x,y} - \Sigma_{(x,y),z} \Sigma_z^{-1} \Sigma_{z,(x,y)}. \quad (12.15)$$

Es ergibt sich also

$$\begin{aligned} \begin{pmatrix} \sigma_{x,x|z}^2 & \sigma_{x,y|z}^2 \\ \sigma_{y,x|z}^2 & \sigma_{y,y|z}^2 \end{pmatrix} &= \begin{pmatrix} 1 & \rho(x, y) \\ \rho(y, x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x, z) \\ \rho(y, z) \end{pmatrix} (1)^{-1} \begin{pmatrix} \rho(x, z) & \rho(y, z) \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho(x, y) \\ \rho(y, x) & 1 \end{pmatrix} - \begin{pmatrix} \rho(x, z)\rho(x, z) & \rho(x, z)\rho(y, z) \\ \rho(y, z)\rho(x, z) & \rho(y, z)\rho(y, z) \end{pmatrix} \\ &= \begin{pmatrix} 1 - \rho(x, z)^2 & \rho(x, y) - \rho(x, z)\rho(y, z) \\ \rho(y, x) - \rho(y, z)\rho(x, z) & 1 - \rho(y, z)^2 \end{pmatrix}. \end{aligned} \quad (12.16)$$

Es ergibt sich also

$$\rho(x, y|z) = \frac{\sigma_{x,y|z}^2}{\sqrt{\sigma_{x,x|z}^2} \sqrt{\sigma_{y,y|z}^2}} = \frac{\rho(x, y) - \rho(x, z)\rho(y, z)}{\sqrt{1 - \rho(x, z)^2} \sqrt{1 - \rho(y, z)^2}}. \quad (12.17)$$

□

Im Falle des Vorliegens von Realisierungen von x, y, z ergibt sich ein entsprechender Schätzer für $\rho(x, y|z)$ mit den Stichprobenkorrelationen $r_{x,y}, r_{x,z}, r_{y,z}$ dann zu

$$r_{x,y|z} = \frac{r_{x,y} - r_{x,z}r_{y,z}}{\sqrt{(1 - r_{x,z}^2)} \sqrt{(1 - r_{y,z}^2)}} \quad (12.18)$$

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 ,

$$\begin{aligned} x &:= \beta_0^{x,z} + \beta_1^{x,z} z \\ y &:= \beta_0^{y,z} + \beta_1^{y,z} z \end{aligned} \quad (12.19)$$

mit Residualvariablen

$$\begin{aligned} e^{x,z} &:= x - \beta_0^{x,z} - \beta_1^{x,z} z \\ e^{y,z} &:= y - \beta_0^{y,z} - \beta_1^{y,z} z \end{aligned} \quad (12.20)$$

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

$$\rho(x, y \setminus z) := \rho(e^{x,z}, e^{y,z}). \quad (12.21)$$

•

Intuitiv entsprechen in obiger Definition die Zufallsvariable $e^{x,z}$ der Zufallsvariable x , aus der der Einfluss von z “herausgerechnet” wurde, und die Zufallsvariable $e^{y,z}$ 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 y und z sowie zwischen x und z wie in der Definition der partiellen Korrelation. Weiterhin seien

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

- $\hat{\beta}_0^{y,z}, \hat{\beta}_1^{y,z}$ die Ausgleichsgeradenparameter für $\{(y_i, z_i)\}_{i=1,\dots,n}$.

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

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

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

•

Für den Fall, dass x, y, z multivariate normalverteilt sind, gibt folgendes Theorem, auf dessen Beweis wir hier verzichten, 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*

$$\rho(x, y|z) = \rho(x, y \setminus z) \quad (12.22)$$

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 $\rho(x, y \setminus z)$ 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 aus den Residualstichprobenkorrelation wie in Definition 12.3 und einmal basierend auf den paarweisen Stichprobenkorrelationen anhand von Theorem 12.1. Schließlich stellt das **R** Paket `ppcor` mit `pcor()` eine Funktion zur automatisierten Auswertung partieller Stichprobenkorrelationen bereit, auch ihre Anwendung demonstrieren wir untenstehend. Das Resultat ist natürlich in allen drei Fällen identisch.

```

1 # Modellformulierung und Datenrealisierung
2 library(MASS) # Multivariate Normalverteilung
3 set.seed(1) # reproduzierbare Daten
4 S = matrix(c( 1,.5,.9, # Kovarianzmatrixparameter \Sigma
5             .5, 1,.5,
6             .9,.5, 1),nrow=3,byrow=TRUE)
7 n = 1e6 # Anzahl Realisierungen von v := (x,y,z)^T
8 xyz = mvrnorm(n,rep(0,3),S) # Realisierungen von v := (x,y,z)^T
9
10 # Partielle Stichprobenkorrelation als Residualstichprobenkorrelation
11 bars = apply(xyz, 2, mean) # Stichprobenmittel
12 s = apply(xyz, 2, sd) # Stichprobenstandardabweichungen
13 c = cov(xyz) # Stichprobenkovarianzen
14 b_xz1 = c[1,3]/c[3,3] # beta_1 (x,z)
15 b_xz0 = bars[1] - b_xz1*bars[3] # beta_0 (x,z)

```

```

16  b_yz1 = c[2,3]/c[3,3]                # beta_1 (y,z)
17  b_yz0 = bars[2] - b_yz1*bars[3]     # beta_0 (y,z)
18  e_xz = xyz[,1] - b_xz1*xyz[,3] - b_xz0 # Residualwerte e^{x,z}
19  e_yz = xyz[,2] - b_yz1*xyz[,3] - b_yz0 # Residualwerte e^{y,z}
20  pr_e = cor(e_xz,e_yz)               # \rho(x,y\z)
21
22  # Partielle Stichprobenkorrelation aus Stichprobenkorrelationen
23  r = cor(xyz)                         # Stichprobenkorrelationsmatrix
24  pr_r_n = r[1,2]-r[1,3]*r[2,3]        # \rho(x,y\z) Formel Zähler
25  pr_r_d = sqrt((1-r[1,3]^2)*(1-r[2,3]^2)) # \rho(x,y\z) Formel Nenner
26  pr_r = pr_r_n/pr_r_d                 # \rho(x,y\z)
27
28  # partielle Stichprobenkorrelation aus Toolbox
29  library(ppcor)                       # Laden der Toolbox
30  pr_t = ppcor(xyz)                   # \rho(x,y\z), \rho(x,z\y), \rho(y,z\z)
31
32  # Ausgabe
33  cat("r(x,y)                          :", r[1,2],
34      "\nr(x,y/z) aus Residuenkorrelation :", pr_e,
35      "\nr(x,y/z) aus Korrelationen      :", pr_r,
36      "\nr(x,y/z) aus ppcor Paket        :", pr_t$estimate[1,2])

```

```

r(x,y)                : 0.5000031
r(x,y/z) aus Residuenkorrelation : 0.1328479
r(x,y/z) aus Korrelationen      : 0.1328479
r(x,y/z) aus ppcor Paket        : 0.1328479

```

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.

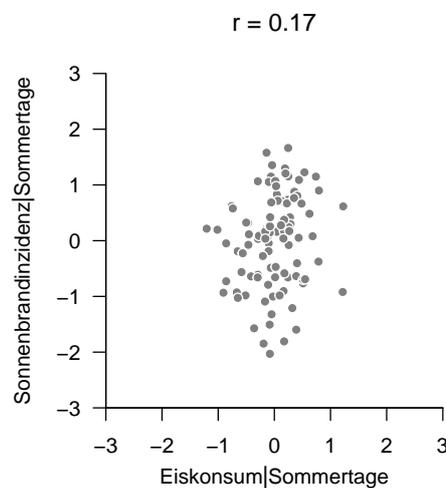


Abbildung 12.3 Evaluation der partiellen Korrelation im Beispielszenario.

Dazu stellt [Abbildung 12.3](#) mit der Achsenbeschriftung **Eiskonsum | Sommertage** die Residualwerte

$$e_i^{x,z} := x_i - \hat{\beta}_0^{x,z} + \hat{\beta}_1^{x,z} z_i \quad (12.23)$$

und mit der Achsenbeschriftung **Sonnenbrandinzidenz | Sommertage** die Residualwerte

$$e_i^{y,z} := y_i - \hat{\beta}_0^{y,z} + \hat{\beta}_1^{y,z} z_i \quad (12.24)$$

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 an.
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 an.
7. Erläutern Sie die Auswertung 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 im Kontext 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*, *Features* oder einfach *Spalten der Designmatrix* genannt. Ähnlich wie bei der einfachen linearen Regression ist das Ziel einer multiplen Regressionsanalyse, das Erklärungspotential für Variationen der abhängigen Variablen durch Variation 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 Wichtungsparameter 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 Pre-Post-Interventions-BDI-Differenzwerte bezeichnen.

Abbildung 13.1 visualisiert den Beispieldatensatz, wobei die dBDI Werte der Höhe gemäß im dreidimensionalen Raum anhand ihrer Koordinaten in der **Alter-Dauer**-Ebene abgetragen sind. Die Visualisierung der multiple Regression, insbesondere für mehr als zwei unabhängige Variablen, stößt sehr schnell an ihre Grenzen.

Tabelle 13.1 Pre-Postinterventions-BDI-Differenzwerte bei Patient:innenalter und Therapiedauer.

Alter	Dauer	dBDI
50	22	22
38	19	23
46	21	18
62	16	7
25	17	22
34	20	24
36	22	32
36	15	20
57	21	18
46	16	8
59	18	11
54	13	4
27	14	17
56	23	21
41	17	17

Alter	Dauer	dBDI
46	21	20
23	22	31
36	23	31
44	20	16
70	18	4

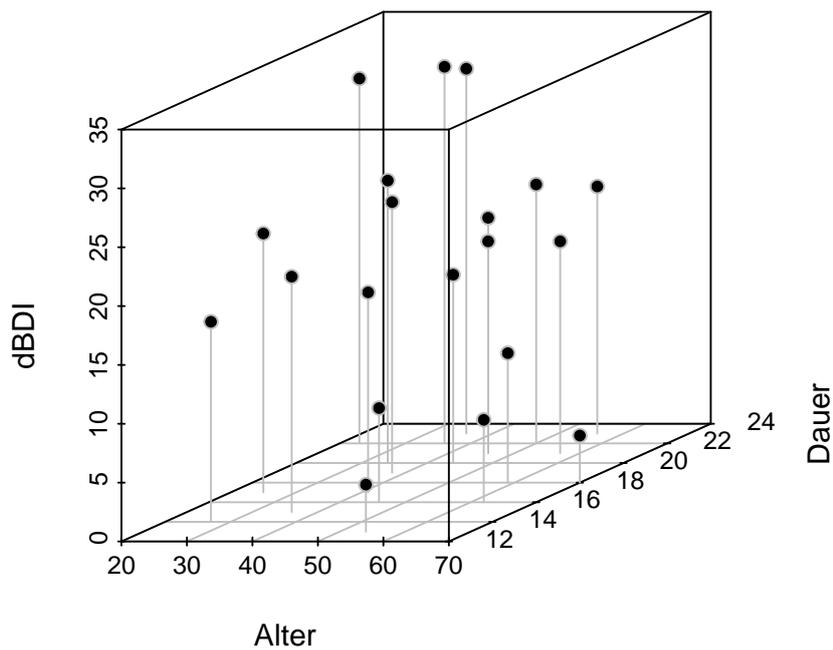


Abbildung 13.1 Visualisierung des Beispieldatensatzes als dreidimensionale Punktwolke

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). v_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

$$v_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \text{ und } \sigma^2 > 0, \quad (13.1)$$

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*,

Kovariaten oder *Features* genannt. Mit

$$x_i := (x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p \text{ und } \beta := (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p \quad (13.2)$$

hat das Modell der multiplen Regression die Datenverteilungsform

$$v_i \sim N(\mu_i, \sigma^2) \text{ u.v. für } i = 1, \dots, n, \text{ wobei } \mu_i := x_i^T \beta. \quad (13.3)$$

In diesem Zusammenhang wird $x_i \in \mathbb{R}^p$ auch als *iter Featurevektor* bezeichnet. Die Designmatrixform des Modells der multiplen Regression schließlich ist gegeben durch

$$v = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (13.4)$$

mit

$$v := (v_1, \dots, v_n)^T, X := (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p} \in \mathbb{R}^{n \times p}, \beta := (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p \text{ und } \sigma^2 > 0. \quad (13.5)$$

•

Folgender **R** Code nutzt das Modell der multiplen Regression auf Definition 13.1, um den in Tabelle 13.1 dargestellten Beispieldatensatz zu erzeugen.

```

1 # Datensimulation
2 library(MASS) # Multivariate Normalverteilung
3 set.seed(10) # reproduzierbare Daten
4 n = 20 # Anzahl Datenpunkte
5 p = 3 # Anzahl Parameter
6 x_1 = round(runif(n,20,80)) # Regressorwerte Alter
7 x_2 = round(runif(n,12,24)) # Regressorwerte Therapiedauer
8 X = matrix(c(rep(1,n),x_1,x_2), nrow = n) # Designmatrix
9 I_n = diag(n) # Identitätsmatrix
10 beta = matrix(c(5,-.5,2), nrow = p) # Betaparametervektor
11 sigsq = 10 # Varianzparameter
12 y = mvrnorm(1, X %*% beta, sigsq*I_n) # eine Realisierung eines n-dimensionalen ZVs
13 y = round(y) # realistische dBDI Werte
14
15 # Dataframeformatierung
16 D = data.frame( # Dataframe Initialisierung
17   Alter = x_1, # Alter
18   Dauer = x_2, # Therapiedauer
19   dBDI = y) # PrePost-BDI Differenzwerte
20
21 # Datenspeicherung
22 fname = file.path(getwd(), "../data/412-Multiple-Regression_Daten.csv")
23 write.csv(D, file = fname, row.names = FALSE)

```

13.3 Modellschätzung

Das Modell der multiplen Regression erlaubt eine allgemeine Einsicht in die Bedeutung des Betaparameterschätzers

$$\hat{\beta} := (X^T X)^{-1} X^T v \quad (13.6)$$

Insbesondere wird bei der Beschäftigung mit dem Betaparameterschätzer im Kontext der multiplen Regression deutlich, dass $X^T v$ die Kovariation der Regressoren mit den Daten und $X^T X$ die Kovariation der Regressoren untereinander quantifiziert, so dass 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*

$$v = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (13.7)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \bar{v} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \\ \frac{r_{v,x_1} - r_{v,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_{v_1}}{s_{x_1}} \\ \frac{r_{v,x_2} - r_{v,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_{v_2}}{s_{x_2}} \end{pmatrix}, \quad (13.8)$$

wobei für die v_i, x_{i1} und x_{i2} mit $i = 1, \dots, n$, $\bar{\cdot}$, s_{\cdot} und $r_{\cdot, \cdot}$ die entsprechenden Stichprobenmittel, Stichprobenstandardabweichungen, und Stichprobenkorrelationen bezeichnen.

Beweis. Wir erinnern zunächst daran, dass die Form des Betaparameterschätzers bekanntlich zum System der Normalgleichungen äquivalent ist (vgl. Kapitel 6)

$$\hat{\beta} = (X^T X)^{-1} X^T v \Leftrightarrow X^T X \hat{\beta} = X^T v. \quad (13.9)$$

Ausschreiben des Normalgleichungssystems für den hier betrachteten ALM Spezialfall ergibt dann zunächst

$$\begin{aligned} X^T X \hat{\beta} &= X^T v \\ \Leftrightarrow \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n v_i \\ \sum_{i=1}^n v_i x_{i1} \\ \sum_{i=1}^n v_i x_{i2} \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n v_i \\ \sum_{i=1}^n v_i x_{i1} \\ \sum_{i=1}^n v_i x_{i2} \end{pmatrix} \end{aligned}$$

und damit

$$\begin{aligned} X^T X \hat{\beta} &= X^T v \\ \Leftrightarrow \begin{pmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} \\ \hat{\beta}_0 \sum_{i=1}^n x_{i2} + \hat{\beta}_1 \sum_{i=1}^n x_{i1} x_{i2} + \hat{\beta}_2 \sum_{i=1}^n x_{i2}^2 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n v_i \\ \sum_{i=1}^n v_i x_{i1} \\ \sum_{i=1}^n v_i x_{i2} \end{pmatrix} \end{aligned}$$

Aus der Gleichung der ersten Vektorkomponenten folgt dann direkt die Form von $\hat{\beta}_0$ mit

$$\begin{aligned} \sum_{i=1}^n v_i &= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} \\ \Leftrightarrow \frac{1}{n} \sum_{i=1}^n v_i &= \hat{\beta}_0 + \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \frac{1}{n} \sum_{i=1}^n x_{i2} \\ \Leftrightarrow \hat{\beta}_0 &= \bar{v} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \end{aligned} \quad (13.10)$$

Einsetzen dieser Form von $\hat{\beta}_0$ in die Gleichung der zweiten Vektorkomponenten ergibt dann

$$\begin{aligned} \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} &= \sum_{i=1}^n v_i x_{i1} \\ (\bar{v} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2) \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} &= \sum_{i=1}^n v_i x_{i1} \\ \bar{v} \sum_{i=1}^n x_{i1} - \hat{\beta}_1 \bar{x}_1 \sum_{i=1}^n x_{i1} - \hat{\beta}_2 \bar{x}_2 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} &= \sum_{i=1}^n v_i x_{i1} \\ \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 - \hat{\beta}_1 \bar{x}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} - \hat{\beta}_2 \bar{x}_2 \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n v_i x_{i1} - \bar{v} \sum_{i=1}^n x_{i1} \\ \hat{\beta}_1 \left(\sum_{i=1}^n x_{i1}^2 - \bar{x}_1 \sum_{i=1}^n x_{i1} \right) + \hat{\beta}_2 \left(\sum_{i=1}^n x_{i1} x_{i2} - \bar{x}_2 \sum_{i=1}^n x_{i1} \right) &= \sum_{i=1}^n v_i x_{i1} - \bar{v} \sum_{i=1}^n x_{i1} \end{aligned}$$

Im Beweis des Theorems zur Ausgleichsgerade (vgl. (1) Regression) haben wir gesehen, dass

$$\begin{aligned} \sum_{i=1}^n x_{i1} x_{i1} - \bar{x}_1 \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1) \\ \sum_{i=1}^n x_{i1} x_{i2} - \bar{x}_2 \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) \\ \sum_{i=1}^n v_i x_{i1} - \bar{v} \sum_{i=1}^n x_{i1} &= \sum_{i=1}^n (v_i - \bar{v})(x_{i1} - \bar{x}_1) \end{aligned} \quad (13.11)$$

Es ergibt sich also, dass

$$\begin{aligned} \hat{\beta}_1 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1) + \hat{\beta}_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) &= \sum_{i=1}^n (v_i - \bar{v})(x_{i1} - \bar{x}_1) \\ \hat{\beta}_1 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1)}{n-1} + \hat{\beta}_2 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{n-1} &= \frac{\sum_{i=1}^n (v_i - \bar{v})(x_{i1} - \bar{x}_1)}{n-1} \end{aligned} \quad (13.12)$$

Mit den Definitionen von Stichprobenstandardabweichung und -korrelation folgt dann weiter

$$\begin{aligned} \hat{\beta}_1 s_{x_1} s_{x_1} + \hat{\beta}_2 c_{x_1, x_2} &= c_{v, x_1} \\ \hat{\beta}_1 \frac{s_{x_1} s_{x_1}}{s_v s_{x_1}} + \hat{\beta}_2 \frac{c_{x_1, x_2}}{s_v s_{x_1}} &= \frac{c_{v, x_1}}{s_v s_{x_1}} \\ \hat{\beta}_1 \frac{s_{x_1}}{s_v} + \hat{\beta}_2 \frac{c_{x_1, x_2}}{s_v s_{x_1}} &= r_{v, x_1} \\ \hat{\beta}_1 \frac{s_{x_1}}{s_v} + \hat{\beta}_2 \frac{c_{x_1, x_2} s_{x_2}}{s_v s_{x_1} s_{x_2}} &= r_{v, x_1} \\ \hat{\beta}_1 \frac{s_{x_1}}{s_v} + \hat{\beta}_2 \frac{s_{x_2}}{s_v} r_{x_1, x_2} &= r_{v, x_1} \end{aligned} \quad (13.13)$$

Definition von

$$b_j := \frac{s_{x_j}}{s_v}, j = 1, 2 \quad (13.14)$$

erlaubt dann die Schreibweise

$$b_1 + b_2 r_{x_1, x_2} = r_{v, x_1}. \quad (13.15)$$

Schließlich folgt analog durch Vertauschen der Subskripte aus der Gleichung der dritten Vektorkomponenten

$$b_1 r_{x_1, x_2} + b_2 = r_{v, x_2} \quad (13.16)$$

Insgesamt haben wir also gesehen, dass die Definition des Betaparameterschätzers im vorliegenden ALM Spezialfall ergibt, dass mit

$$\hat{\beta}_j = b_j \frac{s_v}{s_{x_j}}, j = 1, 2 \quad (13.17)$$

gilt, dass

$$\begin{aligned} r_{v, x_1} &= b_1 + b_2 r_{x_1, x_2} \\ r_{v, x_2} &= b_1 r_{x_1, x_2} + b_2 \end{aligned} \quad (13.18)$$

Damit folgt aus der zweiten Gleichung dann sofort

$$b_2 = r_{v,x_2} - b_1 r_{x_1,x_2}. \quad (13.19)$$

Einsetzen in die erste Gleichung ergibt dann

$$\begin{aligned} b_1 + (r_{v,x_2} - b_1 r_{x_1,x_2}) r_{x_1,x_2} &= r_{v,x_1} \\ \Leftrightarrow b_1 + r_{v,x_2} r_{x_1,x_2} - b_1 r_{x_1,x_2}^2 &= r_{v,x_1} \\ \Leftrightarrow r_{v,x_2} r_{x_1,x_2} + b_1 (1 - r_{x_1,x_2}^2) &= r_{v,x_1} \\ \Leftrightarrow b_1 (1 - r_{x_1,x_2}^2) &= r_{v,x_1} - r_{v,x_2} r_{x_1,x_2} \\ \Leftrightarrow b_1 &= \frac{r_{v,x_1} - r_{v,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \end{aligned} \quad (13.20)$$

Für b_2 ergibt sich damit weiterhin

$$\begin{aligned} b_2 &= r_{v,x_2} - b_1 r_{x_1,x_2} \\ \Leftrightarrow b_2 &= r_{v,x_2} - \left(\frac{r_{v,x_1} - r_{v,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \right) r_{x_1,x_2} \\ \Leftrightarrow b_2 &= \frac{r_{v,x_2} (1 - r_{x_1,x_2}^2)}{1 - r_{x_1,x_2}^2} - \frac{r_{v,x_1} r_{x_1,x_2} - r_{v,x_2} r_{x_1,x_2}^2}{1 - r_{x_1,x_2}^2} \\ \Leftrightarrow b_2 &= \frac{r_{v,x_2} - r_{v,x_2} r_{x_1,x_2}^2 - r_{v,x_1} r_{x_1,x_2} + r_{v,x_2} r_{x_1,x_2}^2}{1 - r_{x_1,x_2}^2} \\ \Leftrightarrow b_2 &= \frac{r_{v,x_2} - r_{v,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \end{aligned} \quad (13.21)$$

Damit folgen dann aber

$$\begin{aligned} \hat{\beta}_1 &= b_1 \frac{s_v}{s_{x_1}} = \left(\frac{r_{v,x_1} - r_{v,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \right) \frac{s_v}{s_{x_1}} \\ \hat{\beta}_2 &= b_2 \frac{s_v}{s_{x_2}} = \left(\frac{r_{v,x_2} - r_{v,x_1} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \right) \frac{s_v}{s_{x_2}} \end{aligned} \quad (13.22)$$

und es ist alles gezeigt. □

Exemplarisch wollen wir die Stichprobenkorrelationsform des Betaparameterschätzers für $\hat{\beta}_1$,

$$\hat{\beta}_1 = \frac{r_{v,x_1} - r_{v,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_v}{s_{x_1}}. \quad (13.23)$$

betrachten. Man erkennt unter anderem:

- Nur im Fall $r_{x_1,x_2} = 0$ und $s_v = s_{x_1}$ gilt $\hat{\beta}_1 = r_{v,x_1}$. 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_{v,x_1} subtrahierte Term $r_{v,x_2} r_{x_1,x_2}$. Die paarweise Korrelation von Regressoren untereinander reduziert also den Betaparameterschätzerwert für einen Regressor.
- Je größer $|r_{v,x_2}|$, desto größer der von r_{v,x_1} subtrahierte Term $r_{v,x_2} r_{x_1,x_2}$. 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_v . Intuitiv erhält also ein Regressor einen höheren Betaparameterschätzerwert, wenn er bei gleich 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.

```

1 # Dateneinlesen
2 fname = file.path(getwd(), "../data/412-Multiple-Regression_Daten.csv")
3 D = read.table(fname, sep = ",", header = TRUE) # Datensatz
4
5 # Modellschätzung
6 y = D$dBDI # Abhängige Variable
7 n = length(y) # Anzahl Datenpunkte
8 X = matrix(c(rep(1,n), D$Alter, D$Dauer), nrow = n) # Designmatrix
9 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
10 eps_hat = y - X %*% beta_hat # Residuenvektor
11 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
12
13 # Betaparameterschätzer aus Stichprobenmittel, -standardabweichungen und -korrelationen
14 y12 = cbind(y, X[, -1]) # y, x_1, x_2 Matrix
15 bars = apply(y12, 2, mean) # Stichprobenmittel
16 s = apply(y12, 2, sd) # Stichprobenstandardabweichungen
17 r = cor(y12) # Stichprobenkorrelationen
18 beta_hat_1 = (r[1,2] - r[1,3]*r[2,3]) / (1 - r[2,3]^2) * (s[1]/s[2]) # \hat{\beta}_1
19 beta_hat_2 = (r[1,3] - r[1,2]*r[2,3]) / (1 - r[2,3]^2) * (s[1]/s[3]) # \hat{\beta}_2
20 beta_hat_0 = bars[1] - beta_hat_1*bars[2] - beta_hat_2*bars[3] # \hat{\beta}_0
21
22 # Ausgabe
23 cat("beta_hat ALM-Schätzer : ", beta_hat,
24     "\nbeta_hat Deskriptivstatistiken : ", c(beta_hat_0, beta_hat_1, beta_hat_2))

```

beta_hat ALM-Schätzer : 4.56351 -0.4388994 1.755541
beta_hat Deskriptivstatistiken : 4.56351 -0.4388994 1.755541

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

$$v = X\beta + \varepsilon, \varepsilon \sim N(0_n, \sigma^2 I_n) \text{ mit } X := \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} \text{ und } \beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (13.24)$$

Dann gilt

$$\hat{\beta} = \begin{pmatrix} \bar{v} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \\ r_{v, x_1 | x_2} \sqrt{\frac{1-r_{v, x_2}^2}{1-r_{x_1, x_2}^2}} \frac{s_v}{s_{x_1}} \\ r_{v, x_2 | x_1} \sqrt{\frac{1-r_{v, x_1}^2}{1-r_{x_2, x_1}^2}} \frac{s_v}{s_{x_2}} \end{pmatrix}, \quad (13.25)$$

wobei für $1 \leq k, l \leq 2$ und $i = 1, \dots, n$

- $r_{v,x_k \setminus x_l}$ die partielle Stichprobenkorrelation der v_i und x_{ik} gegeben die x_{il} ist,
- r_{v,x_k} die Stichprobenkorrelation der v_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

$$\hat{\beta}_1 = \frac{r_{v,x_1} - r_{v,x_2}r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_v}{s_{x_1}} \quad (13.26)$$

Weiterhin haben wir in Kapitel 2 gesehen, dass unter der Annahme der multivariaten Normalverteilung von v, x_1, x_2 ein Schätzer für die partielle Korrelation von v und x_1 gegeben x_2 durch

$$r_{v,2} = \frac{r_{v,x_1} - r_{v,x_2}r_{x_1,x_2}}{\sqrt{1 - r_{v,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \quad (13.27)$$

gegeben ist. Für $\hat{\beta}_1$ ergibt sich somit

$$\begin{aligned} \hat{\beta}_1 &= \frac{r_{v,x_1} - r_{v,x_2}r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_v}{s_{x_1}} \\ &\Leftrightarrow (1 - r_{x_1,x_2}^2) \hat{\beta}_1 = (r_{v,x_1} - r_{v,x_2}r_{x_1,x_2}) \frac{s_v}{s_{x_1}} \\ &\Leftrightarrow \frac{1 - r_{x_1,x_2}^2}{\sqrt{1 - r_{v,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \hat{\beta}_1 = \frac{r_{v,x_1} - r_{v,x_2}r_{x_1,x_2}}{\sqrt{1 - r_{v,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \frac{s_v}{s_{x_1}} \\ &\Leftrightarrow \frac{1 - r_{x_1,x_2}^2}{\sqrt{1 - r_{v,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}} \hat{\beta}_1 = r_{v,x_1 \setminus x_2} \frac{s_v}{s_{x_1}} \end{aligned} \quad (13.28)$$

und damit weiter

$$\begin{aligned} \hat{\beta}_1 &= r_{v,x_1 \setminus x_2} \frac{\sqrt{1 - r_{v,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}}{1 - r_{x_1,x_2}^2} \frac{s_v}{s_{x_1}} \\ &\Leftrightarrow \hat{\beta}_1 = r_{v,x_1 \setminus x_2} \frac{\sqrt{1 - r_{v,x_2}^2} \sqrt{1 - r_{x_1,x_2}^2}}{\left(\sqrt{1 - r_{x_1,x_2}^2}\right)^2} \frac{s_v}{s_{x_1}} \\ &\Leftrightarrow \hat{\beta}_1 = r_{v,x_1 \setminus x_2} \frac{\sqrt{1 - r_{v,x_2}^2}}{\sqrt{1 - r_{x_1,x_2}^2}} \frac{s_v}{s_{x_1}} \\ &\Leftrightarrow \hat{\beta}_1 = r_{v,x_1 \setminus x_2} \sqrt{\frac{1 - r_{v,x_2}^2}{1 - r_{x_1,x_2}^2}} \frac{s_v}{s_{x_1}} \end{aligned} \quad (13.29)$$

□

Im Allgemeinen gilt für $1 \leq i, l \leq k$, dass

$$\hat{\beta}_k \neq r_{v,x_k \setminus x_l}. \quad (13.30)$$

Betaparameterschätzer sind also im Allgemeinen keine partiellen Stichprobenkorrelationen. Allerdings gilt

$$\hat{\beta}_k = r_{v,x_k \setminus x_l} \text{ für } 1 \leq i, l \leq k \quad (13.31)$$

genau dann, wenn

$$s_v = s_{x_1} = s_{x_2} \quad (13.32)$$

und außerdem

$$r_{v,x_k} = r_{x_k,x_l} = 0, \quad (13.33)$$

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

$$|r_{v,x_l}| = |r_{x_k,x_l}|, \quad (13.34)$$

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.

```

1 # Dateneinlesen
2 fname = file.path(getwd(), "../data/412-Multiple-Regression_Daten.csv")
3 D = read.table(fname, sep = ",", header = TRUE) # Datensatz
4
5 # Modellschätzung
6 y = D$dBDI # Abhängige Variable
7 n = length(y) # Anzahl Datenpunkte
8 X = matrix(c(rep(1,n), D$Alter, D$Dauer), nrow = n) # Designmatrix
9 p = ncol(X) # Anzahl Parameter
10 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
11 eps_hat = y - X %*% beta_hat # Residuenvektor
12 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
13
14 # Betaparameterschätzer aus partiellen Korrelationen und Korrelationen
15 library(ppcor) # partielle Korrelationentoolbox
16 y12 = cbind(y,X[,-1]) # y,x_1,x_2 Matrix
17 bars = apply(y12, 2, mean) # Stichprobenmittel
18 s = apply(y12, 2, sd) # Stichprobenstandardabweichungen
19 r = cor(y12) # Stichprobenkorrelationen
20 pr = pcor(y12) # partielle Stichprobenkorrelationen
21 pr = pr$estimate # partielle Stichprobenkorrelationen
22 beta_hat_1 = pr[1,2]*sqrt((1-r[1,3]^2)/(1-r[2,3]^2))*(s[1]/s[2]) # \hat{\beta}_1
23 beta_hat_2 = pr[1,3]*sqrt((1-r[1,2]^2)/(1-r[3,2]^2))*(s[1]/s[3]) # \hat{\beta}_2
24 beta_hat_0 = bars[1] - beta_hat_1*bars[2] - beta_hat_2*bars[3] # \hat{\beta}_0
25
26 # Ausgabe
27 cat("Korrelationen r(y,x_1),r(y,x_2),r(x_1,x_2) : ", c(r[1,2],r[1,3],r[2,3]),
28     "\nPartielle Korrelationen r(y,x_1|x_2), r(y,x_2|x_1) : ", c(pr[1,2],pr[1,3]),
29     "\nbeta_hat ALM Schätzer : ", beta_hat,
30     "\nbeta_hat aus partieller Korrelation : ", c(beta_hat_0,beta_hat_1,beta_hat_2))

Korrelationen r(y,x_1),r(y,x_2),r(x_1,x_2) : -0.7011313 0.6725334 -0.03817849
Partielle Korrelationen r(y,x_1|x_2), r(y,x_2|x_1) : -0.9133605 0.9063215
beta_hat ALM Schätzer : 4.56351 -0.4388994 1.755541
beta_hat aus partieller Korrelation : 4.56351 -0.4388994 1.755541

```

Beispieldatenvisualisierung

Abbildung 13.2 visualisiert den Beispieldatensatz zusammen mit der durch den Betaparameterschätzer definierten Regressionsebene

$$f_{\hat{\beta}} : \mathbb{R}^2 \rightarrow \mathbb{R}, (x_1, x_2) \mapsto f_{\hat{\beta}}(x_1, x_2) := \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \quad (13.35)$$

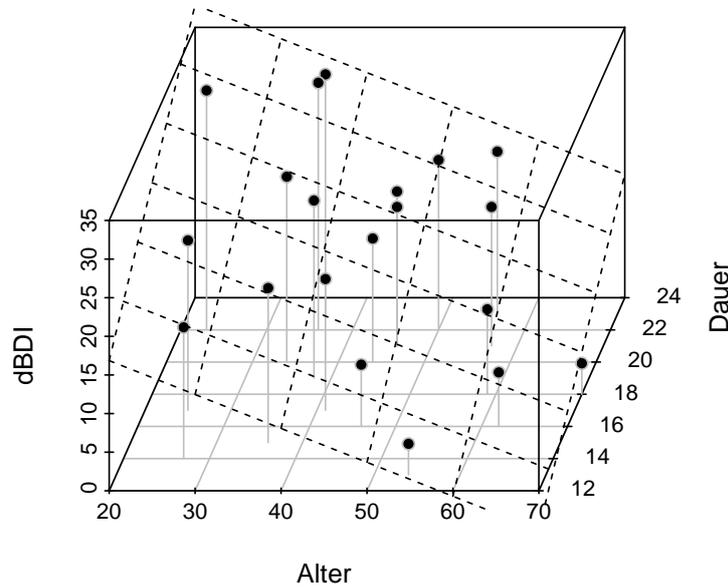


Abbildung 13.2 Visualisierung von Beispieldatensatzes und Regressionsebene

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:

$$c := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Leftrightarrow H_0 : \beta_2 = 0, H_A : \beta_2 \neq 0 \text{ und } c := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \Leftrightarrow H_0 : \beta_3 = 0, H_A : \beta_3 \neq 0 \tag{13.36}$$

Dabei würde das Ablehnen der jeweiligen Nullhypothese jeweils inferenzstatistische Evidenz für einen Effekt des Patient:innenalters bzw. der Therapiedauer auf die Pre-Post-BDI-Differenz 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:

$$c := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, H_0 : \beta_2 - \beta_3 = 0, H_A : \beta_2 - \beta_3 \neq 0. \tag{13.37}$$

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.

```

1 # Dateneinlesen
2 fname = file.path(getwd(), "../_data/412-Multiple-Regression_Daten.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # Modellschätzung
6 y = D$dBDI # Abhängige Variable
7 n = length(y) # Anzahl Datenpunkte
8 X = matrix(c(rep(1,n), D$Alter, D$Dauer), nrow = n) # Designmatrix
9 p = ncol(X) # Anzahl Parameter
10 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
11 eps_hat = y - X %*% beta_hat # Residuenvektor
12 sigsq_hat = (t(eps_hat) %*% eps_hat) / (n-p) # Varianzparameterschätzer
13
14 # Modellevaluation | Parameterinferenz
15 C = cbind(diag(p), matrix(c(0,1,-1), nrow = 3)) # Kontrastgewichtsvektoren
16 ste = rep(NA, ncol(C)) # Kontraststandardfehler
17 tee = rep(NA, ncol(C)) # T-Statistiken
18 pvals = rep(NA, ncol(C)) # p-Werte
19 for(i in 1:ncol(C)){
20   c = C[,i] # Kontrastgewichtsvektor
21   t_num = t(c) %*% beta_hat # Zähler der T-Statistik
22   ste[i] = sqrt(sigsq_hat * t(c) %*% solve(t(X) %*% X) %*% c) # Kontraststandardfehler/Nenner der T-Statistik
23   tee[i] = t_num / ste[i] # T-Statistik
24   pvals[i] = 2 * (1 - pt(abs(tee[i]), n-p)) # p-Wert
25 }
26 # Ausgabe
27 R = data.frame(c(beta_hat, t(C[,4] %*% beta_hat)), ste, tee, pvals)
28 rownames(R) = c("(Interzept)", "Alter", "Dauer", "Alter-Dauer")
29 colnames(R) = c("Estimate", "Std. Error", "t value", "Pr(>|t|)")
30 print(R)

```

	Estimate	Std. Error	t value	Pr(> t)
(Interzept)	4.5635100	4.4102106	1.034760	3.152773e-01
Alter	-0.4388994	0.0474520	-9.249334	4.798503e-08
Dauer	1.7555414	0.1985283	8.842777	9.095118e-08
Alter-Dauer	-2.1944408	0.2023508	-10.844735	4.656992e-09

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 Pre-Post-BDI-Differenzwertvariation, bestätigen. Folgender **R** Code evaluiert den angesprochenen F-Test.

```

1 # Dateneinlesen
2 fname = file.path(getwd(), "../_data/412-Multiple-Regression_Daten.csv")
3 D = read.table(fname, sep = ",", header = TRUE)
4
5 # Modellevaluation
6 y = D$dBDI # Abhängige Variable
7 n = length(y) # Anzahl Datenpunkte
8 X = matrix(c(rep(1,n), D$Alter, D$Dauer), nrow = n) # Designmatrix vollständiges Modell
9 p = ncol(X) # Anzahl Parameter vollständiges Modell
10 p_0 = 1 # Anzahl Parameter reduziertes Modell
11 p_1 = p - p_0 # Anzahl zusätzlicher Parameter im vollst. Modell
12 X_0 = X[,1:p_0] # Designmatrix reduziertes Modell
13 beta_hat_0 = solve(t(X_0) %*% X_0) %*% t(X_0) %*% y # Betaparameterschätzer reduziertes Modell
14 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer vollständiges Modell
15 eps_hat_0 = y - X_0 %*% beta_hat_0 # Residuenvektor reduziertes Modell
16 eps_hat = y - X %*% beta_hat # Residuenvektor vollständiges Modell
17 eh0_eh0 = t(eps_hat_0) %*% eps_hat_0 # RQS reduziertes Modell
18 eh_eh = t(eps_hat) %*% eps_hat # RQS vollständiges Modell
19 sigsq_hat = eh_eh / (n-p) # Varianzparameterschätzer vollst. Modell
20 f = ((eh0_eh0 - eh_eh) / p_1) / sigsq_hat # F-Statistik
21 pval = 1 - pf(f, p_1, n-p) # p-Wert
22
23 # Ausgabe
24 cat("F-statistic:", f, "on", p_1, "and", n-p, "DF", "p-value: ", paste(pval))

```

F-statistic: 85.11912 on 2 and 17 DF p-value: 1.39143241373318e-09

Der Beispieldatensatz enthält also deutliche Evidenz dafür, dass sowohl Patient:innenalter als auch Therapiedauer zur Erklärung der Pre-Post-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.

```
1 fname = file.path(getwd(), "../data/412-Multiple-Regression_Daten.csv") # Datensatzdatei
2 D     = read.table(fname, sep = ",", header = TRUE) # Datensatzeinlesen
3 alm  = lm(dBDI ~ Alter + Dauer, data = D) # Modellformulierung und Modellschätzung
4 summary(alm)
```

Call:
lm(formula = dBDI ~ Alter + Dauer, data = D)

Residuals:

Min	1Q	Median	3Q	Max
-4.4628	-1.4365	0.4511	1.5665	4.9037

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.56351	4.41021	1.035	0.315
Alter	-0.43890	0.04745	-9.249	4.8e-08 ***
Dauer	1.75554	0.19853	8.843	9.1e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.645 on 17 degrees of freedom
Multiple R-squared: 0.9092, Adjusted R-squared: 0.8985
F-statistic: 85.12 on 2 and 17 DF, p-value: 1.391e-09

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 (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, Kovariate und Feature im Rahmen der multiplen Regression.
4. Erläutern Sie, warum $\hat{\beta} \approx \text{Regressorkovariabilität}^{-1} \text{Regressordatenkovariabilität}$ gilt.
5. Erläutern Sie den Zusammenhang zwischen Betaparameterschätzern und Korrelationen in einem multiplen Regressionmodell mit Interzeptprädiktor und zwei kontinuierlichen Prädiktoren anhand der Formel

$$\hat{\beta}_1 = \frac{r_{v,x_1} - r_{v,x_2} r_{x_1,x_2}}{1 - r_{x_1,x_2}^2} \frac{s_v}{s_{x_1}}. \quad (13.38)$$

6. Erläutern Sie den Zusammenhang zwischen Betaparameterschätzern und partieller Korrelation in einem multiplen Regressionmodell mit Interzeptprädiktor und zwei kontinuierlichen Prädiktoren anhand der Formel

$$\hat{\beta}_1 = r_{v,x_1|x_2} \sqrt{\frac{1 - r_{y,x_2}^2}{1 - r_{x_1,x_2}^2}} \frac{s_v}{s_{x_1}}. \quad (13.39)$$

7. $X \in \mathbb{R}^{n \times 2}$ sei die Designmatrix eines multiplen Regressionsmodells mit zwei Prädiktoren und Betaparametervektor $\beta := (\beta_1, \beta_2)^T$. Geben Sie den Kontrastgewichtsvektor an, um die Nullhypothese $H_0 : \beta_1 = \beta_2$ mithilfe der T-Statistik zu testen.

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 & 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>.
- 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, J.L. 1993. „A Simple Argument Showing How to Derive Restricted Maximum Likelihood“.
- Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models*. Third Edition. Los Angeles: SAGE.
- 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 Stimulation*.
- 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>.