

Helmut Strasser

# Lineare Algebra mit statistischen Anwendungen

Einführungsskriptum für Studierende  
von wirtschaftswissenschaftlichen Studienrichtungen

1. Auflage, Wien, 20. Januar 2004



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>7</b>
1.1	Inhalt der Lehrveranstaltung . . . . .	7
1.2	EDV Umgebung . . . . .	8
	Das Arbeitsverzeichnis . . . . .	8
	Libraries . . . . .	9
	Hilfesystem . . . . .	9
<b>2</b>	<b>Vektoren</b>	<b>10</b>
2.1	Grundbegriffe . . . . .	10
	Definition eines Vektors . . . . .	10
	Vektoren in $\mathbb{R}$ . . . . .	10
	Rechenoperationen für Vektoren . . . . .	11
	Interpretation von Vektoren . . . . .	12
2.2	Geometrische Interpretation von Vektoren . . . . .	12
	Punktinterpretation . . . . .	12
	Verschiebungsinterpretation . . . . .	12
	Verschiebung eines Punktes . . . . .	13
	Addition von Verschiebungen . . . . .	14
	Computereperimente . . . . .	15
	Multiplikation einer Verschiebung mit einer Zahl . . . . .	15
	Aufgaben . . . . .	16
2.3	Statistische Interpretation von Vektoren . . . . .	18
2.3.1	Datensätze in $\mathbb{R}$ . . . . .	19
	Das Datenverzeichnis . . . . .	19
	Daten einlesen . . . . .	19
	Variablenamen: Attach . . . . .	20
2.3.2	Univariate Datenanalyse mit $\mathbb{R}$ . . . . .	20
	Sortieren und Häufigkeiten . . . . .	20
	Variablentypen . . . . .	21
	Diagramme . . . . .	22
	Statistische Maßzahlen . . . . .	22
	Aufgaben . . . . .	23
<b>3</b>	<b>Matrizen</b>	<b>24</b>
3.1	Matrizenrechnung mit $\mathbb{R}$ . . . . .	24
	Der Begriff einer Matrix . . . . .	24
	Erzeugung von Matrizen . . . . .	26
	Matrizenrechnung . . . . .	27
	Elimination und inverse Matrizen . . . . .	29
	Apply und Sweep . . . . .	31
	Aufgaben . . . . .	32

3.2	Input-Output Analyse . . . . .	35
3.2.1	Die Kostenseite des Input-Output Modells . . . . .	35
3.2.2	Ein Musterbeispiel . . . . .	37
<b>4</b>	<b>Metrische Konzepte</b>	<b>40</b>
4.1	Skalares Produkt und Norm . . . . .	40
	Skalares Produkt . . . . .	40
	Skalarprodukt in $\mathbb{R}$ . . . . .	41
	Norm . . . . .	41
	Norm von Vektoren in $\mathbb{R}$ . . . . .	42
	Einheitsvektoren und Normierung . . . . .	42
	Orthogonale Vektoren . . . . .	43
	Beispiele in $\mathbb{R}$ . . . . .	44
	Aufgaben . . . . .	44
4.2	Orthogonale Projektion . . . . .	45
	Das LSQ-Problem . . . . .	45
	Orthogonale Zerlegung . . . . .	47
	Der Winkel zwischen Vektoren . . . . .	48
	Aufgaben . . . . .	49
4.3	Statistische Anwendungen . . . . .	51
4.3.1	Der Mittelwert . . . . .	51
	Hinweise zu $\mathbb{R}$ . . . . .	51
4.3.2	Die Varianz . . . . .	52
	Hinweise zu $\mathbb{R}$ . . . . .	53
4.3.3	Kovarianz und Korrelation . . . . .	53
	Hinweise zu $\mathbb{R}$ . . . . .	54
	Aufgaben . . . . .	54
<b>5</b>	<b>Die Methode der kleinsten Quadrate</b>	<b>55</b>
5.1	Das einfache Regressionsmodell . . . . .	55
5.1.1	Modellanpassung (model fitting) . . . . .	55
	LSQ mit $\mathbb{R}$ . . . . .	57
5.1.2	Die mathematische Lösung . . . . .	58
	Das Absolutglied . . . . .	58
	LSQ Optimierung als orthogonale Projektion . . . . .	59
	Die Normalgleichungen . . . . .	60
5.1.3	Beurteilung der Modellanpassung . . . . .	61
	Das Ausmaß des Erklärungsbeitrags . . . . .	62
	Das Signifikanzproblem . . . . .	62
	Das Relevanzproblem . . . . .	65
5.1.4	Aufgaben . . . . .	68
5.2	Das zweifache Regressionsmodell . . . . .	69
5.2.1	Einführung . . . . .	69
	LSQ mit $\mathbb{R}$ . . . . .	70
5.2.2	Die mathematische Lösung . . . . .	71
	Die Normalgleichungen . . . . .	71
5.2.3	Die Beurteilung der Modellanpassung . . . . .	73
	RSquare . . . . .	73
	ANOVA-Zerlegungen . . . . .	74
	Modellselektion . . . . .	78
	Signifikanzprüfung der Parameter . . . . .	81
	Durchführung der Modellselektion . . . . .	83

5.2.4	Aufgaben . . . . .	85
5.3	Ein qualitativer Prädiktor . . . . .	86
5.3.1	Die Fragestellung . . . . .	86
5.3.2	Prädiktoren mit zwei Kategorien . . . . .	86
	Modell und Modellvergleich . . . . .	86
	Parametrisierung durch Haupteffekte . . . . .	89
5.3.3	Prädiktoren mit mehr als zwei Kategorien . . . . .	92
	Parametrisierung durch Haupteffekte . . . . .	92
	Parametrisierung mit gewählten Kontrasten . . . . .	95
5.3.4	Aufgaben . . . . .	98
5.4	Additive Modelle und Wechselwirkungen . . . . .	100
5.4.1	Zwei quantitative Prädiktoren . . . . .	100
5.4.2	Ein qualitativer und ein quantitativer Prädiktor . . . . .	102
	Das additive Modell . . . . .	102
	Das Modell mit Wechselwirkung . . . . .	106
5.4.3	Zwei qualitative Variable . . . . .	108
	Die Problemstellung . . . . .	108
	Additive Modelle . . . . .	110
	Modelle mit Wechselwirkung . . . . .	113
	Orthogonale Versuchspläne . . . . .	115
5.4.4	Aufgaben . . . . .	115
<b>6</b>	<b>Quadratische Funktionen</b>	<b>117</b>
6.1	Funktionen von mehreren Variablen . . . . .	117
6.1.1	Grundlegende Definitionen . . . . .	117
	Lineare Funktionen . . . . .	117
	Quadratische Funktionen . . . . .	117
6.1.2	Graphische Darstellungen . . . . .	118
	Funktionsgraphen . . . . .	118
	Niveauliniendiagramm . . . . .	120
	Schnittliniendiagramm . . . . .	120
6.1.3	Aufgaben . . . . .	121
6.2	Die Hauptachsentransformation . . . . .	122
6.2.1	Orthogonale Matrizen . . . . .	122
	Orthogonale 2x2-Matrizen . . . . .	124
6.2.2	Diagonalisierung von symmetrischen Matrizen . . . . .	125
	Numerische Diagonalisierung . . . . .	125
6.2.3	Anwendung auf quadratische Funktionen . . . . .	127
	Einführung . . . . .	127
	Der allgemeine Fall . . . . .	129
6.2.4	Aufgaben . . . . .	129
6.3	Konvexe und konkave Funktionen . . . . .	130
6.3.1	Grundlegende Definitionen . . . . .	130
6.3.2	Quadratische Funktionen . . . . .	131
6.3.3	Aufgaben . . . . .	132
<b>7</b>	<b>Quadratische Optimierung</b>	<b>135</b>
7.1	Grundlagen aus der Analysis . . . . .	135
7.1.1	Einleitung . . . . .	135
7.1.2	Richtungsableitung . . . . .	135
7.1.3	Partielle Ableitungen . . . . .	137
7.1.4	Der Gradient . . . . .	138

	Gradientenfelder . . . . .	141
7.1.5	Aufgaben . . . . .	141
7.2	Globale Optimierung . . . . .	141
7.2.1	Kritische Punkte . . . . .	141
7.2.2	Homogene quadratische Funktionen . . . . .	143
7.2.3	Inhomogene quadratische Funktionen . . . . .	143
7.2.4	Aufgaben . . . . .	145
7.3	Optimierung mit Nebenbedingungen . . . . .	146
7.3.1	Optimierung unter einer linearen Nebenbedingung . . . . .	147
	Die Methode von Lagrange . . . . .	148
7.3.2	Optimierung unter mehreren linearen Nebenbedingungen . . . . .	151
	Die Methode von Lagrange . . . . .	152
7.3.3	Das Basislemma . . . . .	154
7.3.4	Aufgaben . . . . .	155
7.4	Portfolio-Optimierung . . . . .	156
7.4.1	Grundlagen . . . . .	156
	Der Datensatz . . . . .	156
	Erste Schritte . . . . .	157
	Portfolios . . . . .	158
7.4.2	Optimale Portfolios . . . . .	160
	Vorbemerkung . . . . .	160
	Minimierung der Varianz . . . . .	160
	Minimierung der Varianz bei vorgegebener Rendite . . . . .	162
	Maximierung der Rendite . . . . .	164
7.4.3	Aufgaben . . . . .	165

# 1

## Einleitung

### 1.1 Inhalt der Lehrveranstaltung

Der Inhalt der LV ist Mathematik. Die Mathematik besteht aus Denkstrategien und Denkinstrumenten, deren intuitiver Hintergrund einerseits in der Raum- und Zeitvorstellung und andererseits in der Vorstellung von Manipulationen (anordnen, vertauschen, abzählen, gegenüberstellen, usw.) liegt. An der Mathematik sind vor allem zwei Dinge erstaunlich. Erstens ist die Mathematik auch dann erfolgreich anwendbar, wenn der Anwendungsbereich nichts oder nur wenig zu tun hat mit Vorgängen in Raum und Zeit oder mit Manipulationen. Zweitens läßt sich das Gebäude der Mathematik weitgehend logisch zwingend und widerspruchsfrei errichten.

Wenn man Mathematik mit dem Ziel lernen will, angewandte sozial- und wirtschaftswissenschaftliche Probleme damit zu bearbeiten, dann muß man trotzdem mit dem intuitiven Hintergrund vertraut werden, der genetisch ursprünglich hinter den mathematischen Konzepten steckt. Außerdem ist es notwendig, den logisch deduktiven Aufbau bis zu einem gewissen Grad zu beherrschen.

In dieser LV wird Lineare Algebra („Vektorrechnung“) behandelt. Es werden Begriffe und ihr intuitiver Hintergrund, Fakten und ihre deduktive Grundlage, sowie Methoden und ihr algorithmischer Ablauf vorgestellt. Die Auswahl der Themen erfolgt nach den Bedürfnissen des sozial- und wirtschaftswissenschaftlichen Wissenschaftsbetriebs.

Die Lineare Algebra ist ein Teilgebiet der Mathematik, dessen Ursprünge vor allem aus zwei Quellen stammen: Die Geometrie von Punkten und Verschiebungen, und das Lösen von linearen Gleichungen. Bei der mathematischen Durchleuchtung dieser Problembereiche sind im Laufe der Jahrhunderte Begriffe und Lösungsansätze entstanden, die für die gesamte Mathematik und ihre Anwendungen von grundlegender Bedeutung sind.

Die Begriffswelt der Linearen Algebra hat sich mittlerweile vollständig von den intuitiven Konzepten der Geometrie und von den Anlässen der Gleichungslehre gelöst. Trotzdem sind Konzepte aus Geometrie und Gleichungslehre beim pädagogischen Aufbau des Gebäudes unerlässlich. Auch in fortgeschrittenem Stadium muß die Problemlösung bei einem Anwendungsfall oft unter Benützung intuitiver geometrischer Konzepte gefunden werden.

## 1.2 EDV Umgebung

Wir werden im Laufe der LV viele mathematische Methoden auf dem Computer realisieren. Das Programmpaket R ermöglicht uns die interaktive Benützung von grundlegenden Operationen in Form fest programmierter Makrobefehle.

An dieser Stelle verweisen wir zunächst auf die Homepage von R:

`http://www.r-project.org`

Von dieser Webseite aus kann R kostenlos geladen und installiert werden. Das selbstinstallierende File

`http://www.ci.tuwien.ac.at/R/bin/windows/base/SetupR.exe`

ist 15.5 MB groß.

Das offizielle Einführungsskript ist:

`http://www.ci.tuwien.ac.at/R/doc/manuals/R-intro.pdf`

Man installiert R nach den Anweisungen, die mitgeliefert werden. Anschließend startet man R, indem man

`RGui.exe`

aufruft.

### Das Arbeitsverzeichnis

---

R arbeitet immer in einem bestimmten Arbeitsverzeichnis (`workdir`), welches mit einem Menüpunkt gesetzt werden kann. Wir erstellen in Windows z.B. ein Verzeichnis mit dem Namen:

`C:\Workspace\R`

Um das Arbeitsverzeichnis über ein Kommando festzulegen, geben wir ein:

```
> setwd("C:/Workspace/R")
```

Die Kontrolle ergibt:

```
> getwd()
[1] "C:\\Workspace\\R"
```

Soll das Arbeitsverzeichnis schon beim Start von R festgelegt werden, dann muß dieses Kommando im File **RProfile** enthalten sein. Dieses File befindet sich im R-Installationsverzeichnis unter **etc**.

Um den Inhalt des Arbeitsverzeichnisses anzuzeigen, verwenden wir:

```
> dir()
...
```

## Libraries

---

R kann durch benutzerdefinierte Funktionen erweitert werden. Zur Unterstützung dieser LV verwenden wir benutzerdefinierte Funktionen aus der Library **LinAlg.r**. Diese Library steht auf der Homepage der LV zur Verfügung. Um sie verwenden zu können, muß sie im Arbeitsverzeichnis von R abgespeichert werden und dann aktiviert werden:

```
> source("LinAlg.r")
```

## Hilfesystem

---

R besitzt ein ausführliches Online-Hilfesystem. Information über das Hilfesystem erhält man durch

```
help("help")
```

## 2

## Vektoren

## 2.1 Grundbegriffe

**Definition eines Vektors**

(2.1) DEFINITION:

Unter einem **Vektor** versteht man eine Zahlenliste. Die Zahlen, aus denen der Vektor besteht, nennt man seine **Komponenten**. Die Anzahl der Komponenten ist die **Länge** des Vektors.

Man kann Vektoren schriftlich als Spaltenvektoren oder als Zeilenvektoren anschreiben. Ein Spaltenvektor der Länge  $n$  hat die allgemeine Form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{oder} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{usw..}$$

Zeilenvektoren haben die Form

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad \text{oder} \quad \mathbf{a} = (a_1, a_2, \dots, a_n) \quad \text{usw..}$$

Später wird die Unterscheidung zwischen Spalten- und Zeilenvektoren eine große Rolle spielen. Das ist aber in diesem Kapitel noch nicht der Fall. Wenn wir von Vektoren sprechen, ist es im Moment gleichgültig, ob wir Spaltenvektoren oder Zeilenvektoren meinen. Es ist nur wichtig, daß wir bei einem bestimmten Typ bleiben. Alles was wir in diesem Kapitel sagen, gilt für Spaltenvektoren und für Zeilenvektoren in gleicher Weise.

Für die Menge aller Vektoren einer festen Länge  $n$  verwendet man das Symbol  $\mathbb{R}^n$ .

**Vektoren in R**

In R legt man Vektoren durch Aufzählung ihrer Komponenten fest.

```
> a=c(1,2,3,4,5)
> a
[1] 1 2 3 4 5
> length(a)
[1] 5
```

Einzelne Komponenten erreicht man durch Indizierung:

```
> a[3]
[1] 3
```

Es gibt noch andere Möglichkeiten, Vektoren zu erzeugen. Die folgenden Beispiele sind selbsterklärend. Weitere Informationen finden sich in der Hilfefunktion von R.

```
> rep(3,5)
[1] 3 3 3 3 3
> rep(c(1,2),3)
[1] 1 2 1 2 1 2
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
> -2:4
[1] -2 -1 0 1 2 3 4
> seq(from=3,to=9,by=2)
[1] 3 5 7 9
```

Für die interaktive Eingabe von Vektoren kann man die R-Funktionen **scan** und **edit** verwenden.

Manchmal werden wir Vektoren erzeugen, die aus Zufallszahlen bestehen, z.B.

```
> x=rnorm(100)
> x
[1] -0.047329181 -0.185997151 -1.275701124 -0.119879542
[5] -0.031689555 0.022956186 1.370558024 1.695163348
[9] 1.049487133 -1.817979639 -1.360895121 0.406407390
[13] -0.235913334 0.906614895 1.235141218 0.620106328
[17] 1.010912207 -0.221201378 -1.283650520 -1.416084392
[21] 1.050887909 -0.875688734 -0.241006555 -1.209303901
[25] -1.373914288 -1.677828735 -1.797911162 0.459355755
[29] 0.077614778 0.663994912 0.114210723 1.150920234
...
```

## Rechenoperationen für Vektoren

---

Wir wissen, wie man mit Vektoren rechnet. Die Addition und die Multiplikation mit einer Zahl erfolgen komponentenweise (siehe Skriptum Mathematik).

Sehen wir uns diese Vorgänge in R an:

```
> a=c(1,2,3)
> b=c(5,6,7)
> a+b
[1] 6 8 10
> a-b
[1] -4 -4 -4
> 3*a-2*b
[1] -7 -6 -5
```

## Interpretation von Vektoren

---

Vektoren (Zahlenlisten) sind formale Objekte. Solche formalen Objekte haben Interpretationen. Bei den Interpretationen muß man zwischen angewandten und intuitiven Interpretationen unterscheiden. Es gibt unzählige angewandte Interpretationen von Vektoren in den Wirtschaftswissenschaften. Man denke etwa an Bedarfsvektoren, Inputvektoren, Outputvektoren usw. (vgl. Skriptum). In der Statistik werden Vektoren als Datenlisten interpretiert. Alle diese Interpretationen sind nicht geometrisch. Die geometrische Interpretation von Vektoren (Punkte und Verschiebungen) dagegen ist keine angewandte Interpretation, da dabei Vektoren durch idealisierte Objekte unserer Vorstellung interpretiert werden. Die geometrische Interpretation ist Teil des intuitiven Hintergrundes der Vektoralgebra.

## 2.2 Geometrische Interpretation von Vektoren

Wir erklären nun die geometrische Interpretation von Vektoren im  $\mathbb{R}^2$ . Leider ist dies insofern kompliziert, als es (mindestens) zwei geometrische Objekte gibt, deren mathematische Fassung ein Vektor ist.

### Punktinterpretation

---

Wir denken uns die Zeichenebene als Menge von Punkten. Wenn wir in der Zeichenebene ein **Koordinatensystem** einführen, dann können wir jedem Punkt der Zeichenebene einen Vektor der Dimension 2 zuordnen, dessen Komponenten genau die Koordinaten des Punktes sind. Auf diese Weise identifizieren wir die Zeichenebene mit  $\mathbb{R}^2$ , der Menge aller Vektoren der Dimension 2. (Abbildung 2.1)

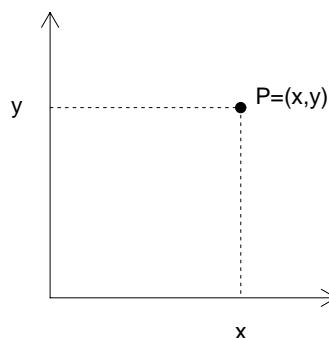


Abbildung 2.1: Punktinterpretation von Vektoren

### Verschiebungsinterpretation

---

Wir denken uns wieder die Zeichenebene als Menge von Punkten mit einem Koordinatensystem. Allerdings sind es jetzt nicht die Punkte, die wir durch Vektoren darstellen wollen. Es sind vielmehr **Verschiebungen** (Translationen), auf die wir jetzt unser Augenmerk richten. Unter einer Verschiebung verstehen wir eine geradlinige Bewegung in einer bestimmten

Richtung um eine bestimmte Strecke. Geometrisch können wir eine Verschiebung durch einen Pfeil veranschaulichen. Es muß aber dabei klar sein, daß der Anfangspunkt des Pfeiles irgendwo in der Zeichenebene fixiert werden kann. Gleichgültig wo das ist, es wird durch alle Pfeile einer bestimmten Richtung und einer bestimmten Länge immer die gleiche Verschiebung symbolisiert.

Wie ordnen wir nun einer solchen, durch einen Pfeil symbolisierten Verschiebung einen Vektor zu? Wir projizieren einen (beliebigen) Verschiebungspfeil auf die Achsen und messen die Längen der Projektionen (mit Vorzeichen + bei gleicher und - bei entgegengesetzter Richtung zu den Achsen). Der Vektor (Zahlenpaar), dessen Komponenten diese Meßergebnisse sind, ist der Verschiebungsvektor. (Abbildung 2.2)

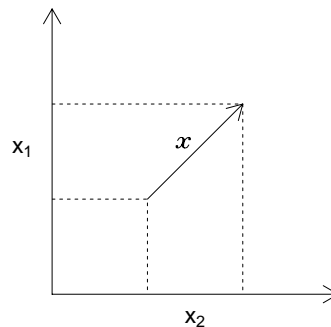


Abbildung 2.2: Verschiebungsinterpretation von Vektoren

Wir haben nun zwei unterschiedliche intuitive Interpretationen von Vektoren im  $\mathbb{R}^2$  kennengelernt. Eine intuitive Interpretation eines Formalismus ist natürlich völlig wertlos, wenn es nicht weitere Entsprechungen (Isomorphismen) gibt, die beim Interpretationsvorgang gültig sind. In der Vektorrechnung sind das die elementaren **Vektoroperationen**.

Unter den elementaren Vektoroperationen versteht man die **Addition von Vektoren** und die **Multiplikation eines Vektors mit einer Zahl**. Wir interessieren uns nun dafür, durch welche geometrischen Operationen zwischen Punkten und Verschiebungen die Rechenoperationen zwischen Vektoren veranschaulicht werden können.

## Verschiebung eines Punktes

Es seien  $P$  und  $Q$  zwei Punkte der Zeichenebene und es sei  $\overrightarrow{PQ}$  jene Verschiebung der Punkte der Zeichenebene, die den Punkt  $P$  in den Punkt  $Q$  überführt. Wir veranschaulichen diese Verschiebung durch den Pfeil von  $P$  nach  $Q$ . Es seien  $\mathbf{a}$  und  $\mathbf{b}$  die Vektoren der Punkte  $P$  und  $Q$ , und  $\mathbf{x}$  sei der Verschiebungsvektor  $\overrightarrow{PQ}$ . Dann ist aus der Zeichung (Abbildung 2.3) unmittelbar ablesbar, daß

$$\mathbf{x} = \mathbf{q} - \mathbf{p} \quad \text{bzw.} \quad \mathbf{p} + \mathbf{x} = \mathbf{q}.$$

In Worten heißt das: Anfangspunkt + Verschiebungsvektor = Endpunkt, oder: Verschiebungsvektor = Endpunkt – Anfangspunkt. Die Vektoraddition kann also interpretiert werden als Verschiebung eines Punktes in einen anderen Punkt.

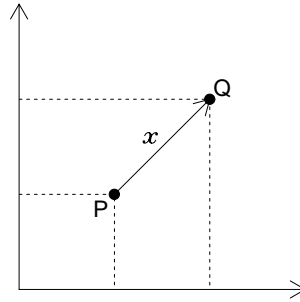


Abbildung 2.3: Verschiebung von Punkten

## Addition von Verschiebungen

Es seien  $\mathbf{a}$  und  $\mathbf{b}$  zwei Verschiebungsvektoren. Wir wollen jene Verschiebung finden, deren Verschiebungsvektor die Summe  $\mathbf{c} := \mathbf{a} + \mathbf{b}$  ist. Zu diesem Zweck wählen wir irgendeinen beliebigen Punkt  $P$  der Zeichenebene und hängen daran den Pfeil, der den Vektor  $\mathbf{a}$  besitzt. Den Endpunkt des Pfeiles nennen wir  $Q$ . An diesen Punkt  $Q$  hängen wir den Pfeil, der den Vektor  $\mathbf{b}$  besitzt. Den Endpunkt dieses zweiten Pfeiles nennen wir  $R$ . Aus der entstehenden Zeichnung (Abbildung 4) können wir folgendes ablesen: Wenn wir die Verschiebungen  $\mathbf{a}$  und  $\mathbf{b}$  hintereinander ausführen, so entsteht insgesamt eine neue Verschiebung, die durch den Pfeil  $\overrightarrow{PR}$  veranschaulicht werden kann. Zu diesem Pfeil gehört der Vektor  $\mathbf{c} = \mathbf{a} + \mathbf{b}$ . Folglich entspricht die Addition von Vektoren der Hintereinanderausführung (Verkettung) von Verschiebungen.

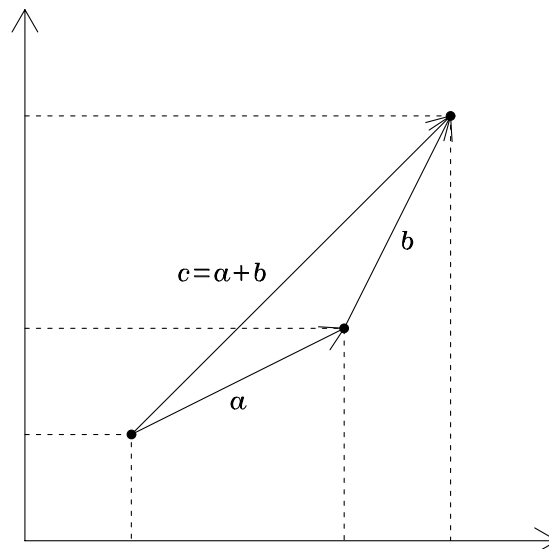


Abbildung 2.4: Vektoraddition

Wir haben damit zwei völlig unterschiedliche intuitive Interpretationen der Vektoraddition kennengelernt. Der Umgang mit beiden Möglichkeiten ist durch Erfahrung erlernbar.

Für die Vektoraddition gilt das **Kommutativgesetz**  $a + b = b + a$ . So selbstverständlich dies scheinen mag, ist es doch in der Verschiebungsinterpretation eine erstaunliche Aussage: Geht man 3 Schritte nach rechts und anschließend 5 Schritte nach vorne, so landet man am gleichen Punkt, wie wenn man zunächst 5 Schritte nach vorne und erst dann 3 Schritte nach rechts geht. Wenn Sie das erstaunlich finden, dann haben Sie das Kommutativgesetz der Vektoraddition wirklich verstanden. Andernfalls ist es Ihnen noch nicht gelungen, sich die in Ihrer Frühkindheit von Ihnen spielerisch entdeckten mathematischen Theoreme ausreichend bewußt zu machen.

## Computorexperimente

---

Um die geometrische Interpretation von Vektoren besser zu verstehen, führen wir einige Computorexperimente durch. Wir verwenden dabei die Funktionen **coord**, **point** und **shift** aus der Library **LinAlg.r**.

Wir zeichnen zunächst ein Koordinatensystem:

```
> coord(xlim=c(-5,5),ylim=c(-5,5))
```

Mit den Funktionen **point** und **shift** können wir Punkte und Pfeile einzeichnen. Nach jedem Zeichenvorgang muß mit einem Mausklick die Position der Beschriftung festgelegt werden. Die Funktion **shift** besitzt außerdem eine Option **from=c(0,0)**, mit der der Anfangspunkt eines Pfeils festgelegt werden kann.

Mit diesen Funktionen können wir nun experimentieren.

```
> a=c(2,3)
> point(a)
> shift(a)
> shift(a, from=c(1,-1))
```

Die Verschiebung eines Punktes:

```
> p=c(-2,-3)
> q=c(1,1)
> point(p)
> point(q)
> shift(q-p, from=p)
```

Die Vektoraddition:

```
> p=c(-3,-3)
> a=c(1,3)
> b=c(4,2)
> shift(a, from=p)
> shift(b, from=p+a)
> shift(a+b, from=p)
```

## Multiplikation einer Verschiebung mit einer Zahl

---

Bei der Multiplikation eines Vektors mit einer positiven Zahl behält der Pfeil, der den Vektor darstellt, seine Richtung bei. Es ändert sich seine Länge. Wir werden später sehen, wie man die Länge eines Vektors berechnet.

Eine negative Zahl dreht bei Multiplikation mit einem Vektor die Pfeilrichtung um.

Aus diesen Beobachtungen ergeben sich die folgenden Begriffe.

(2.2) DEFINITION:

Zwei Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  heißen **proportional** (in Zeichen  $\mathbf{a} \parallel \mathbf{b}$ ), wenn entweder  $\mathbf{a} = \lambda \mathbf{b}$  oder  $\mathbf{b} = \mu \mathbf{a}$  für  $\lambda \in \mathbb{R}$  und  $\mu \in \mathbb{R}$ . Sind zwei proportionale Vektoren  $\mathbf{a}, \mathbf{b}$  von Null verschieden, so heißen sie **gleich gerichtet**, wenn der Proportionalitätsfaktor positiv ist, und **entgegengesetzt gerichtet**, wenn der Proportionalitätsfaktor negativ ist.

## Aufgaben

(2.3) **Aufgabe:** Gegeben seien die zwei Vektoren  $\mathbf{a} = (1, 4, -3, 2, 5, -1)$  und  $\mathbf{b} = (16, 13, 10, 7, 4, 1)$ . Berechnen Sie

1.  $\mathbf{a} + \frac{1}{2}\mathbf{b}$ ,
2.  $2\mathbf{a} - \mathbf{b}$ .

(2.4) **Aufgabe:** Gegeben sind die Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  aus Beispiel 2.3, der Vektor  $\mathbf{c} = (a_1, a_2, a_3, a_1, a_2, a_3)$ , der aus Komponenten von  $\mathbf{a}$  gebildet wird, und der Vektor  $\mathbf{d} = (4, 5, 6, 7, 8, 9)$ . Berechnen Sie  $3\mathbf{c} + 2\mathbf{d} - \mathbf{a} - \frac{1}{2}\mathbf{b}$ .

(2.5) **Aufgabe:** Bei der Zusammenstellung eines Wertpapierportfolios haben Sie die Auswahl unter Anteilsscheinen an drei Investmentfonds  $F_1, F_2$  und  $F_3$ . Alle drei Fonds haben ihr Vermögen ausschließlich in Aktien der Unternehmen  $A, B$  und  $C$  investiert mit der folgenden prozentuellen Verteilung:

Unternehmen	Anteile in %		
	$F_1$	$F_2$	$F_3$
$A$	10	50	20
$B$	70	40	20
$C$	20	10	60

Wenn Sie 25 Stück von  $F_1$ , 120 Stück von  $F_2$  und 60 Stück von  $F_3$  erwerben, wie verteilt sich dann Ihr investiertes Kapital auf die Unternehmen  $A, B$  und  $C$ ?

**Lösung:** Wir formen drei Vektoren  $\mathbf{f}_1, \mathbf{f}_2$  und  $\mathbf{f}_3$  mit den Anteilsverteilungen der Fonds:

- >  $\mathbf{f}_1 = \mathbf{c}(0.1, 0.7, 0.2)$
- >  $\mathbf{f}_2 = \mathbf{c}(0.5, 0.4, 0.1)$
- >  $\mathbf{f}_3 = \mathbf{c}(0.2, 0.2, 0.6)$

Ein Portfolio zusammenzustellen bedeutet, eine **Linearkombination** dieser drei Vektoren zu bilden:

- >  $\mathbf{z} = 25 * \mathbf{f}_1 + 120 * \mathbf{f}_2 + 60 * \mathbf{f}_3$
- >  $\mathbf{z}$
- [1] 74.5 77.5 53.0

Um die prozentuelle Verteilung des Portfolios auf die drei Unternehmen zu erhalten, berechnen wir die Summe der Komponenten und dividieren  $\mathbf{z}$  durch diese Summe:

- >  $\mathbf{v} = \mathbf{z} / \text{sum}(\mathbf{z})$
- >  $\mathbf{v}$
- [1] 0.3634146 0.3780488 0.2585366

In diesem Portfolio sind daher 36.3% des Kapitals in  $A$ , 37.8% in  $B$  und 25.9% in  $C$  investiert.

(2.6) **Aufgabe:** Es seien  $\mathbf{a} = (12, 43, 37, -32)$ ,  $\mathbf{b} = (-22, 42, 0, 25)$ ,  $\mathbf{c} = (-31, 62, 84, -90)$  Vektoren.

- Bestimme einen Vektor  $\mathbf{d}$  so, daß  $3(\mathbf{a} + 5\mathbf{d}) = -\mathbf{b} + \mathbf{d}$ .
- Bestimme einen Vektor  $\mathbf{d}$  so, daß  $\mathbf{a} + 2\mathbf{d} - 5\mathbf{b} + 7\mathbf{d} = 2(\mathbf{d} + 3\mathbf{b}) - \mathbf{c}$ .
- Gibt es eine Zahl  $s$ , sodaß  $(3 + 2s)\mathbf{a} + 4\mathbf{b} = s\mathbf{a} - 2\mathbf{c}$ ?
- Gibt es einen Vektor  $\mathbf{q}$ , sodaß  $3\mathbf{a} + 2\mathbf{q} + 4\mathbf{b} = \mathbf{q} - 2\mathbf{c}$ ?

(2.7) **Aufgabe:** Es seien  $A = (1, 1)$ ,  $B = (4, 2)$  und  $C = (2, 5)$  die Eckpunkte eines Dreiecks.

- Wie lauten die Verschiebungsvektoren der Seiten des Dreiecks?
- Wie kann man den Verschiebungsvektor einer Seite durch die Verschiebungsvektoren der beiden anderen Seiten berechnen?
- Illustrieren Sie die Aufgabe durch eine Grafik.

(2.8) **Aufgabe:** Gegeben sind die Punkte  $A = (3, 4)$  und  $B = (5, 2)$ .  $\mathbf{a}$  und  $\mathbf{b}$  seien die dazugehörigen Vektoren.

- Berechnen Sie den zum Punkt  $C$  gehörenden Vektor  $\mathbf{c} = \frac{1}{2}(\mathbf{a} + \mathbf{b})$ .
- Illustrieren Sie die Aufgabe durch eine Grafik.

Unter einer **Schwerlinie** eines Dreiecks versteht man die Verbindungsstrecke von einem Eckpunkt zum Halbierungspunkt der gegenüberliegenden Seite.

(2.9) **Aufgabe:** Es seien  $A = (1, 1)$ ,  $B = (4, 2)$  und  $C = (2, 5)$  die Eckpunkte eines Dreiecks.

- Berechnen Sie die Verschiebungsvektoren der Schwerlinien des Dreiecks.
- Wie läßt sich eine Schwerlinie durch die beiden anderen berechnen?
- Illustrieren Sie die Aufgabe durch eine Grafik.

Unter einem Parallelogramm versteht man ein Viereck, bei dem gegenüberliegenden Seiten die gleichen Verschiebungsvektoren besitzen (dh. gleich lang und parallel sind).

(2.10) **Aufgabe:** Es seien  $A = (1, 1)$ ,  $B = (4, 2)$  und  $C = (3, 4)$  und  $D = (2, 3)$  Eckpunkte eines Vierecks.

- Berechnen Sie die Halbierungspunkte der Seiten des Vierecks.
- Zeigen Sie, daß die Halbierungspunkte ein Parallelogramm aufspannen.
- Illustrieren Sie die Aufgabe durch eine Grafik.

(2.11) **Aufgabe:** Es seien  $\mathbf{u} = (2, 1)$  und  $\mathbf{v} = (2, 3)$  zwei Verschiebungsvektoren.

- Zeichnen Sie die  $\mathbf{u}$  und  $\mathbf{v}$  ausgehend von Ursprung des Koordinatensystems.
- Berechnen Sie und zeichnen Sie den Verschiebungsvektor  $\mathbf{v} - \mathbf{u}$ .
- Es sei  $\mathbf{a} = 3\mathbf{u}$  und  $\mathbf{b} = 3\mathbf{v}$ . Berechnen und zeichnen Sie  $\mathbf{a}$ ,  $\mathbf{b}$  und  $\mathbf{b} - \mathbf{a}$ .
- Begründen Sie rechnerisch, dass  $\mathbf{b} - \mathbf{a}$  zu  $\mathbf{v} - \mathbf{u}$  parallel ist.

Es seien  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  beliebige Vektoren. Dann nennt man die Menge

$$M := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \alpha\mathbf{a} + \beta\mathbf{b} \text{ für } \alpha + \beta = 1, \alpha \geq 0, \beta \geq 0\}$$

die **Verbindungsstrecke** zwischen den Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ .

(2.12) **Aufgabe:** Gegeben seien die Vektoren  $\mathbf{a} = (-3, 1)$  und  $\mathbf{b} = (2, -3)$ . Geben Sie drei Punkte an, die auf der Verbindungsstrecke zwischen  $\mathbf{a}$  und  $\mathbf{b}$  liegen. Illustrieren Sie die Aufgabe durch eine Grafik.

Es seien  $\mathbf{c} \in \mathbb{R}^n$  und  $\mathbf{t} \in \mathbb{R}^n$ ,  $\mathbf{t} \neq \mathbf{o}$ , Vektoren. Dann heißt die Menge

$$g(\mathbf{a}, \mathbf{t}) := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{a} + \lambda\mathbf{t}, \lambda \in \mathbb{R}\}$$

eine **Gerade** durch den Punkt  $a$  mit dem Richtungsvektor  $t$ .

(2.13) **Aufgabe:** Gegeben sei die Gerade  $g$  durch den Punkt  $P = (2, 1)$  mit Richtungsvektor  $t = (3, 4)$ .

a) Zeichnen Sie fünf verschiedene Punkte der Geraden  $g$ .

b) Welche der folgenden Vektoren kommen als Richtungsvektoren der Geraden  $g$  nicht in Frage:  $(1, \frac{4}{3})$ ,  $(-\frac{3}{2}, -2)$ ,  $(3, -4)$ ,  $(0, -\frac{5}{3})$

#### (2.14) TESTFRAGEN

1. Welche geometrischen Interpretationen von Vektoren kennen Sie ?
2. Erklären Sie die geometrische Interpretation der elementaren rechenoperationen für Vektoren (Addition, Subtraktion, Multiplikation mit einer Zahl).
3. Wie berechnet man den Halbierungspunkt der Verbindungsstrecke von zwei Punkten ?

## 2.3 Statistische Interpretation von Vektoren

In der Statistik spielen Vektoren eine große Rolle. Sie treten aber in mehreren unterschiedlichen Bedeutungen auf. Die wichtigsten davon sind: **Datenlisten** und **Datenvektoren**. Worin besteht nun der Unterschied zwischen diesen Begriffen.

Zunächst erinnern wir uns daran, daß man in der Statistik grundsätzlich zwischen **Untersuchungsobjekten (Cases)** und **Merkmalen (Variablen)** unterscheidet. Das Datenmaterial, welches einer statistischen Untersuchung zugrundeliegt, besteht in der Regel aus einer **Datenmatrix**, deren Zeilen den Cases und deren Spalten den Variablen entsprechen. Ein einzelner Spaltenvektor enthält also die **Ausprägungen (realizations)** einer einzelnen Variablen für alle Cases. So etwas nennt man eine **Datenliste**.

**In der Statistik werden Datenlisten einzelner Variabler durch Spaltenvektoren dargestellt.**

Wenn man sich aber für die Ausprägungen mehrerer Variabler bei einem einzigen Case interessiert, muß man eine Zeile der Datenmatrix anschauen. so etwas nennt man einen **Datenvektor**.

**In der Statistik werden Datenvektoren einzelner Cases durch Zeilenvektoren dargestellt.**

Die Unterscheidung zwischen Datenlisten und Datenvektoren ist uns aus der elementaren Statistik geläufig. Untersucht man z.B. die Korrelation von zwei Variablen, so kann man das Datenmaterial entweder als Liste von Datenpaaren (zweidimensionale Datenvektoren) ansehen und diese Datenpaare in einem Streudiagramm zeichnen. Oder aber man sieht das Datenmaterial als Paar von Datenlisten an und versucht eine Liste durch die andere zu erklären.

Wir beginnen damit, statistische Datenlisten (also die Spalten einer Datenmatrix, oder in anderen Worten: statistische Variable) mit den Begriffen und Methoden der Vektorechnung zu behandeln. Später werden wir auch die Zeilen von mehrspaltigen Datenmatrizen mit Mitteln der linearen Algebra untersuchen.

Wir wissen aus der LV Statistik bereits sehr viel über Datenlisten. Daher können wir uns hier darauf beschränken, die technischen Fragen zu besprechen.

## 2.3.1 Datensätze in R

Statistische Daten werden in R meistens als Datenframes gespeichert. Dabei handelt es sich um Datenmatrizen mit Zeilennamen und Spaltennamen. Darüber hinaus können Spalten, die qualitative Variable enthalten, das Klassenattribut **factor** erhalten. Sie werden dann vom R-System anders behandelt als Spalten mit quantitativen Variablen.

### Das Datenverzeichnis

---

Wir erstellen ein Unterverzeichnis unseres Arbeitsverzeichnisses, z.B.

```
C:\Workspace\R\data
```

In das Unterverzeichnis **data** kopieren wir die Datei:

```
statlab.txt
```

Die Datei **statlab.txt** enthält den Datensatz, den wir für unsere statistischen Beispiele verwenden werden. Der Datensatz und seine Beschreibung stehen auf der Homepage der LV zu Verfügung.

### Daten einlesen

---

Die einfachste Möglichkeit besteht darin, den Befehl **data(statlab)** zu verwenden. Um das zu tun, muß sich das File **statlab.txt** allerdings im Unterverzeichnis **data** des Arbeitsverzeichnisses befinden.

Es ist aber hilfreich zu wissen, wie man ein Textfile grundsätzlich einlesen kann. Dazu dienen die folgenden Hinweise.

Das File **statlab.txt** enthält Daten im Ascii-Format mit Spaltenüberschriften. In jeder Zeile sind die Daten durch Blanks getrennt. Dezimaltrennzeichen ist der Punkt.

Um den Datensatz **statlab** mit dem folgenden Befehl einlesen zu können, muß sich das File **statlab.txt** im Arbeitsverzeichnis befinden.

```
> statlab <- read.table(file="statlab.txt",header=TRUE,sep=" ",dec=".")
```

Wichtig: `sep=""` bedeutet, daß jede zusammenhängende Folge von Blanks als ein einziges Trennzeichen behandelt wird. Würden wir `sep=" "` eingeben, dann würde jedes einzelne Blank als Trennzeichen interpretiert.

Wir stellen nun die Anzahl der Zeilen und Spalten des Datensatzes fest:

```
> dim(statlab)
[1] 1296  34
```

sowie die Variablennamen:

```
> names(statlab)
 [1] "CODE"   "CBSEX"  "CBB"    "CBLGTH" "CBWGT"  "CBMO"
 [7] "CBD"    "CBHR"   "CTHGHT" "CTWGT"  "CTL"    "CTPEA"
[13] "CTRA"   "MBB"    "MBAG"   "MBWGT"  "MBO"    "MBSM"
[19] "MTHGHT" "MTWGT"  "MTE"    "MTO"    "MTSM"   "FBAG"
[25] "FBO"    "FBSM"   "FTHGHT" "FTWGT"  "FTE"    "FTO"
```

```
[31] "FTSM" "FIB" "FIT" "FC"
```

Um den Datensatz zu Gänze anzuzeigen, verwenden wir

```
> data.entry(statlab)
```

Dadurch wird der Datensatz in einem Spreadsheet-ähnlichen Fenster angezeigt.

Wichtig: Wenn wir Daten bei dieser Anzeige ändern, hat das unmittelbar Auswirkung auf den gespeicherten Datensatz. Er wird dadurch verändert.

## Variablenamen: Attach

---

Nach dem Laden eines Datensatzes sind die einzelnen Variablen noch nicht unter ihrem Namen erreichbar:

```
> CBB
Error: Object "CBB" not found
```

Dies liegt daran, daß sich der Datensatz **statlab** noch nicht im Suchpfad befindet:

```
> search()
[1] ".GlobalEnv" "package:ctest" "Autoloads"
[4] "package:base"
```

Um den Zugriff auf die einzelnen Variablen des Datensatzes zu vereinfachen, verwenden wir:

```
> attach(statlab)
```

Dieser Befehl fügt dem Suchpfad das Objekt **statlab** hinzu:

```
> search()
[1] ".GlobalEnv" "statlab" "package:ctest" "Autoloads"
[5] "package:base"
```

und nun sind alle Variablen des Datensatzes unter ihrem Spaltennamen dem System bekannt und erreichbar:

```
> CBB
  [1] 2 2 7 6 5 5 7 6 5 6 2 5 5 5 5 1 5 5 6 5 5 8 6
  ...
```

## 2.3.2 Univariate Datenanalyse mit R

Wir haben nun statistische Variable als Vektoren gespeichert und in R verfügbar. Nun besprechen wir die einfachsten statistischen Analyseschritte.

### Sortieren und Häufigkeiten

---

Bei Datenvektoren, die Variable darstellen, interessiert man sich in der Regel nur für die Verteilung der Komponenten, dh. man vernachlässigt die Reihenfolge.

Daher besteht ein erster Schritt der Datenanalyse im Sortieren der Daten:

```
> x=c(3,1,4,7)
> sort(x)
[1] 1 3 4 7
```

Den Rang (die Platzziffer) der einzelnen Komponenten erhält man durch:

```
> order(x)
[1] 2 1 3 4
> x[order(x)]
[1] 1 3 4 7
```

Falls Bindungen auftreten, ist die Menge der auftretenden Werte von Interesse:

```
> MTE
  [1] 1 3 2 2 2 2 2 2 3 2 1 2 1 2 2 2 2 2 3 0 2 2 3 3 2 2 2 2 2 3 2
 [32] 2 2 2 3 2 2 2 2 2 2 2 3 3 2 2 3 3 1 2 2 2 2 3 2 2 3 3 3 2 2 2
 [63] 4 3 3 3 3 3 3 2 2 2 2 2 3 2 2 3 2 1 2 2 3 3 1 2 2 2 2 3 3 3 3
 [94] 3 3 2 1 2 2 3 2 2 2 3 3 1 2 3 2 3 2 2 3 3 4 2 3 3 3 2 3 2 2 3
  ...
> unique(MTE)
[1] 1 3 2 0 4
Levels: 0 1 2 3 4
```

Ein Häufigkeitstabelle erhalten wir durch

```
> table(MTE)
MTE
 0    1    2    3    4
26 101 494 384 291
```

## Variablentypen

---

Für statistische Analysen ist grundsätzlich der Variablentyp zu beachten. Die allereinfachste Unterscheidung erfolgt zwischen qualitativen und quantitativen Variablen. In R werden qualitative Variablen als Faktoren bezeichnet.

Beim Einlesen einer Datenmatrix in einen Dataframe betrachtet R (als Voreinstellung) jede numerische Datenspalte als quantitativ, und jede nichtnumerische Datenspalte als Faktor.

Wir wollen die Variable CBB untersuchen. Das ist eine numerisch kodierte qualitative Variable. R glaubt zunächst (auf Grund der numerischen Kodierung), daß es sich um eine quantitative Variable handelt. Das kann man unter anderem daraus sehen, daß mit dem Befehl

```
> summary(CBB)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   5.000   6.000   5.531   6.000   9.000
```

eine Zusammenfassung der Datenliste erstellt wird, die sich für qualitative Variable nicht eignet.

Wenn numerisch kodierte Datenspalten als Faktoren behandelt werden sollen, dann muß das dem R-System mitgeteilt werden:

```
> CBB <- factor(CBB)
> summary(CBB)
```

1	2	3	4	5	6	7	8	9
61	64	22	6	481	376	145	56	85

Diese Häufigkeitstabelle ist nun eine angemessene Zusammenfassung der Datenliste einer qualitativen Variablen.

## Diagramme

---

Einfache Diagramme erhält man mit der Funktion **plot**. Diese Funktion liefert für Faktoren (qualitative Variable) die üblichen Balkendiagramme. Bei quantitativen Variablen gibt es aber keine brauchbare Voreinstellung.

Wir gehen daher so vor:

```
> plot(unique(CTHGHT), table(CTHGHT), type="h", lwd=1)
```

Ein interessantes anderes Bild bietet

```
> x=rnorm(100)
> plot(unique(x), table(x), type="h", lwd=1)
```

Um eine empirische Verteilungsfunktion zu zeichnen, laden wir die Library **stepfun**. Die Herstellung des Diagramms erfolgt durch

```
> library(stepfun)
> plot(ecdf(CTWGT))
```

Es sind in R noch zahlreiche weitere Diagramme verfügbar (Histogramme, Dichteschätzer, Boxplots, usw.). Wir verweisen auf das Hilfesystem von R und weiterführende Literatur.

## Statistische Maßzahlen

---

R stellt Funktionen zur Berechnung aller wichtigen Verteilungsmaßzahlen zu Verfügung.

Mittelwert:	<b>mean</b>
Varianz:	<b>var</b>
Standardabweichung:	<b>sd</b>
Median:	<b>median</b>
Quantile:	<b>quantile</b>
Maximum:	<b>max</b>
Minimum:	<b>min</b>

Bei Varianz und Standardabweichung ist zu beachten, dass es sich dabei um jene Versionen handelt, bei denen durch  $n - 1$  statt durch  $n$  dividiert wird.

Man kann diese Maßzahlen auch auf die Spalten einer Datenmatrix anwenden:

```
> xx=data.frame(CBLGTH, MBAG, FIT)
> apply(xx, 2, mean)
      CBLGTH      MBAG      FIT
20.39252 28.27778 155.19444
> apply(xx, 2, median)
CBLGTH  MBAG  FIT
 20.5   27.0 144.0
> apply(xx, 2, quantile, prob=0.1)
```

CBLGTH	MBAG	FIT
19	21	87

## Aufgaben

---

(2.15) **Aufgabe:** Beurteilen Sie im Datensatz **statlab** die Typen der Variablen.

(2.16) **Aufgabe:** Bestimmen Sie die Häufigkeitsverteilungen der Variablen im Datensatz **statlab**.

(2.17) **Aufgabe:** Datensatz **statlab**: Wie hoch ist das durchschnittliche, das minimale und das maximale Geburtsgewicht von Kindern mit einer Mutter, die älter als 35 Jahre ist ?

**Lösung:**

```
> i=which(MBAG>35)
> mean(CBWGT[i])
[1] 7.552632
> max(CBWGT[i])
[1] 10.6
> min(CBWGT[i])
[1] 4.5
```

(2.18) **Aufgabe:** Datensatz **statlab**: Wie häufig sind Elternpaare, bei denen der Vater jünger als die Mutter ist ?

**Lösung:**

```
> sum(MBAG>FBAG)
[1] 123
```

(2.19) **Aufgabe:** Datensatz **statlab**: Sind unter den Familien mit den 10 Prozent intelligentesten Kindern die Mütter über 30 Jahren überrepräsentiert oder unterrepräsentiert ?

**Lösung:**

```
> x=quantile(CTPEA,0.9)
> x
90%
 91
> i=which(MBAG>30)
> sum(CTPEA[i]>=91)/length(i)
[1] 0.1442080
```

## 3

## Matrizen

## 3.1 Matrizenrechnung mit R

Wir haben in der LV Mathematik schon viel über Matrizen und Matrizenrechnung erfahren. Daher genügt es vorerst, darüber zu sprechen, wie man in R mit Matrizen arbeitet.

### Der Begriff einer Matrix

---

Im Gegensatz zu einem Vektor hat eine Matrix nicht nur eine Länge. Sie hat vielmehr zwei Dimensionen, nämlich die Anzahl der Zeilen und die Anzahl der Spalten:

```
> A=matrix(1:6,nrow=2,ncol=3)
> A
      [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

Das war ganz einfach. Wir wollen aber genauer hinter die Kulissen schauen.

Ein Matrix in R hat das Attribut **dim**, welches die Anzahl der Zeilen und Spalten enthält:

```
> dim(A)
[1] 2 3
```

R weiss was eine Matrix ist:

```
> is.matrix(A)
[1] TRUE
> is.vector(A)
[1] FALSE
```

Wenn wir das **dim**-Attribut verändern, dann ändert sich die Gestalt der Matrix, aber nicht der Wert ihrer Komponenten:

```
> dim(A)=c(3,2)
> A
      [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
> dim(A)=c(1,6)
> A
```

```

      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    2    3    4    5    6
> dim(A)=c(6,1)
> A
      [,1]
[1,]    1
[2,]    2
[3,]    3
[4,]    4
[5,]    5
[6,]    6

```

Man kann das **dim**-Attribute auch ganz streichen. Dann wird die Matrix zu einem Vektor:

```

> dim(A)=NULL
> A
[1] 1 2 3 4 5 6
> is.matrix(A)
[1] FALSE
> is.vector(A)
[1] TRUE

```

Umgekehrt kann man aus jedem Vektor eine Matrix machen, indem man das **dim**-Attribut setzt:

```

> y=1:4
> y
[1] 1 2 3 4
> dim(y)=c(2,2)
> y
      [,1] [,2]
[1,]    1    3
[2,]    2    4

```

Spaltenvektoren und Zeilenvektoren im Sinn der Matrizenrechnung sind nicht Vektoren im Sinn von R, sondern spezielle Matrizen:

```

> a=1:5
> a
[1] 1 2 3 4 5
> is.vector(a)
[1] TRUE
> dim(a)=c(1,5)
> a
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    2    3    4    5
> is.vector(a)
[1] FALSE
> is.matrix(a)
[1] TRUE

```

## Erzeugung von Matrizen

---

Im einfachsten Fall werden Matrizen dadurch erzeugt, dass man einen Vektor erzeugt, und diesen mit einem **dim**-Attribut versieht. Die Funktion **matrix** führt genau das durch.

Matrizen können neben dem **dim**-Attribut auch noch das Attribut **dimnames** haben. Das sind Namen für die einzelnen Zeilen und Spalten:

```
> A=matrix(1:6,nrow=2,ncol=3)
> A
      [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
> dimnames(A)=list(c("a","b"),c("x","y","z"))
> A
   x y z
a 1 3 5
b 2 4 6
```

Diese Namen stehen in einer Liste. Komponenten einer Liste kann man einzeln ansprechen:

```
> dimnames(A)[[1]]
[1] "a" "b"
> dimnames(A)[[2]]
[1] "x" "y" "z"
```

Eine andere Möglichkeit, Matrizen zu erzeugen, besteht im Zusammenbinden von Vektoren. Entweder als Spalten:

```
> x=c(1,2,3)
> y=c(4,5,6)
> cbind(x,y)
   x y
[1,] 1 4
[2,] 2 5
[3,] 3 6
```

oder als Zeilen:

```
> rbind(x,y)
  [,1] [,2] [,3]
x    1    2    3
y    4    5    6
```

Schließlich gibt es einfache Möglichkeit, Diagonalmatrizen zu erzeugen:

```
> diag(c(1,2,3))
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    2    0
[3,]    0    0    3
```

zum Beispiel die Einheitsmatrix:

```
> diag(rep(1,8))
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    0    0    0    0    0    0    0
```

```
[2,] 0 1 0 0 0 0 0 0
[3,] 0 0 1 0 0 0 0 0
[4,] 0 0 0 1 0 0 0 0
[5,] 0 0 0 0 1 0 0 0
[6,] 0 0 0 0 0 1 0 0
[7,] 0 0 0 0 0 0 1 0
[8,] 0 0 0 0 0 0 0 1
```

## Matrizenrechnung

---

Die elementaren Rechenoperationen funktionieren so, wie man es erwartet:

```
> A=matrix(1:12,nrow=4,ncol=3)
> B=matrix(5:16,nrow=4,ncol=3)
> A+B
      [,1] [,2] [,3]
[1,]   6  14  22
[2,]   8  16  24
[3,]  10  18  26
[4,]  12  20  28
> A-B
      [,1] [,2] [,3]
[1,]  -4  -4  -4
[2,]  -4  -4  -4
[3,]  -4  -4  -4
[4,]  -4  -4  -4
> 2*A
      [,1] [,2] [,3]
[1,]   2  10  18
[2,]   4  12  20
[3,]   6  14  22
[4,]   8  16  24
> -B
      [,1] [,2] [,3]
[1,]  -5  -9 -13
[2,]  -6 -10 -14
[3,]  -7 -11 -15
[4,]  -8 -12 -16
```

Man kann Matrizen mit gleichen **dim**-Attributen auch komponentenweise multiplizieren und dividieren:

```
> A*B
      [,1] [,2] [,3]
[1,]   5  45 117
[2,]  12  60 140
[3,]  21  77 165
[4,]  32  96 192
> A/B
      [,1]      [,2]      [,3]
[1,] 0.2000000 0.5555556 0.6923077
[2,] 0.3333333 0.6000000 0.7142857
```

```
[3,] 0.4285714 0.6363636 0.7333333
[4,] 0.5000000 0.6666667 0.7500000
```

Allerdings haben diese Rechenoperationen nichts mit der Matrixmultiplikation im Sinn der Matrizenrechnung zu tun. Die eigentliche Matrixmultiplikation erfolgt so:

```
> C=matrix(3:8,nrow=3,ncol=2)
> A**%C
      [,1] [,2]
[1,]   68  113
[2,]   80  134
[3,]   92  155
[4,]  104  176
```

Die Matrixmultiplikation erfordert, dass die Matrizen so zusammenpassen, wie es die Matrizenrechnung vorschreibt:

```
> A**%B
Error in A ** B : non-conformable arguments
```

Matrizen kann man transponieren:

```
> A
      [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
> t(A)
      [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    5    6    7    8
[3,]    9   10   11   12
```

Beim Transponieren eines Produkts ändert sich die Reihenfolge der Faktoren:

```
> A
      [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
> C
      [,1] [,2]
[1,]    3    6
[2,]    4    7
[3,]    5    8
> t(A**%C)
      [,1] [,2] [,3] [,4]
[1,]   68   80   92  104
[2,]  113  134  155  176
> t(C)**%t(A)
      [,1] [,2] [,3] [,4]
[1,]   68   80   92  104
[2,]  113  134  155  176
```

## Elimination und inverse Matrizen

---

Wir haben das Eliminationsverfahren kennengelernt, und wir wissen, dass man damit lineare Gleichungssystem lösen und Matrizen invertieren kann.

Das Eliminationsverfahren verwandelt eine Matrix in eine Form, die wir als Stufenform bezeichnet haben. Der Fachausdruck dafür lautet **row reduced echelon form**, kurz **rref**, und das bedeutet: durch Zeilenelimination gewonnene Skelettform.

In R ist **rref** als Funktion nicht vorgesehen, und daher haben wir diese Funktion in unsere Library aufgenommen:

```
> A=matrix(rnorm(20),4,5)
> rref(A)
$echelon.form
      [,1] [,2] [,3] [,4]      [,5]
[1,]    1    0    0    0 -2.15298825
[2,]    0    1    0    0  0.27724896
[3,]    0    0    1    0  0.02313954
[4,]    0    0    0    1  0.80876803

$row.operations
      [,1]      [,2]      [,3]      [,4]
[1,]  0.30450910 -2.1179512  0.7316129 -0.61341499
[2,] -0.43476917 -0.1767498 -0.1754193 -0.03315625
[3,] -0.08645671 -0.8147642  0.1773673  0.12281785
[4,]  0.14879617  0.3670251 -0.5871976  0.22482689
```

Ein anderes Beispiel wird später noch eine wichtige Rolle spielen:

```
> B=matrix(21:29,3,3)
> rref(B)
$echelon.form
      [,1] [,2] [,3]
[1,]    1    0   -1
[2,]    0    1    2
[3,]    0    0    0

$row.operations
      [,1] [,2] [,3]
[1,] -8.333333  8    0
[2,]  7.333333 -7    0
[3,]  1.000000 -2    1
```

Die Funktion **rref** gibt eine Liste zurück, die aus den beiden Komponenten **echelon.form** und **row.operations** besteht. Die zweite Komponenten enthält jene Matrix, die die Zeilenoperationen angibt, die man zur Berechnung der ersten Komponenten, eben der Skelettform durchführen muss:

```
> L=rref(B)$row.operations
> L
      [,1] [,2] [,3]
[1,] -8.333333  8    0
[2,]  7.333333 -7    0
[3,]  1.000000 -2    1
```

```
> L%%B
      [,1] [,2] [,3]
[1,]    1    0   -1
[2,]    0    1    2
[3,]    0    0    0
```

Bei einer regulären quadratischen Matrix besteht die Skelettfom aus der Einheitsmatrix. Auf diese Weise führt das Eliminationsverfahren zur inversen Matrix:

```
> k=sample(-5:5,9,replace=TRUE)
> k
[1] -3  3  5  3  0  5 -2  2  3
> A=matrix(k,3,3)
> A
      [,1] [,2] [,3]
[1,]   -3    3   -2
[2,]    3    0    2
[3,]    5    5    3
> rref(A)
$echelon.form
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

```
$row.operations
      [,1]      [,2] [,3]
[1,] -3.3333333 -6.3333333  2
[2,]  0.3333333  0.3333333  0
[3,]  5.0000000 10.0000000 -3
```

```
> C=rref(A)$row.operations
> C%%A
      [,1] [,2]      [,3]
[1,] 1.000000e+00  0 8.881784e-16
[2,] -5.551115e-17  1 0.000000e+00
[3,] 0.000000e+00  0 1.000000e+00
> round(C%%A,5)
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Reguläre Matrizen kann man in R mit **solve** invertieren:

```
> C=matrix(c(1,4,3,7,5,2,4,8,6),nrow=3,ncol=3)
> C
      [,1] [,2] [,3]
[1,]    1    7    4
[2,]    4    5    8
[3,]    3    2    6
> D=solve(C)
> D
```

```

      [,1]      [,2]      [,3]
[1,] -1.0  2.4285714 -2.5714286
[2,]  0.0  0.4285714 -0.5714286
[3,]  0.5 -1.3571429  1.6428571
> C%*%D

```

```

      [,1]      [,2] [,3]
[1,]    1 0.000000e+00  0
[2,]    0 1.000000e+00  0
[3,]    0 4.440892e-16  1
> round(C%*%D,5)

```

```

      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1

```

Man sollte sich aber darüber im klaren sein, dass das Invertieren ein Vorgang ist, der nicht immer klappen muss. Meistens merkt es R, wenn eine Matrix nicht invertierbar ist:

```

> C=matrix(1:9,3,3)
> solve(C)
Error in solve.default(C) : Lapack routine dgesv:
system is exactly singular

```

manchmal aber auch nicht:

```

> C=matrix(21:29,3,3)
> C
      [,1] [,2] [,3]
[1,]  21   24   27
[2,]  22   25   28
[3,]  23   26   29
> D=solve(C)
> D
      [,1]      [,2]      [,3]
[1,] -3.886798e+14  7.773596e+14 -3.886798e+14
[2,]  7.773596e+14 -1.554719e+15  7.773596e+14
[3,] -3.886798e+14  7.773596e+14 -3.886798e+14
> C%*%D

```

```

      [,1] [,2] [,3]
[1,] -0.375  1.625 -1.125
[2,]  0.500  0.500 -0.500
[3,] -0.625 -0.625  0.125

```

Man sollte sich daher immer vorher vergewissern, ob eine Matrix tatsächlich invertierbar ist. Man macht dies, indem man ihren Rang feststellt, z.B. mit dem Eliminationsverfahren.

## Apply und Sweep

---

Die Funktion **apply** berechnet Zusammenfassungen von Zeilen oder Spalten einer Matrix.

Das folgende Beispiel berechnet Zeilensummen und Spaltenmittelwerte:

```
> A=matrix(1:12,4,3)
> apply(A,1,sum)
[1] 15 18 21 24
> apply(A,2,mean)
[1] 2.5 6.5 10.5
```

Man kann aber auch spezielle Funktionen definieren, um sie auf Zeilen oder Spalten anzuwenden, z.B. Quadratsummen der Zeilen:

```
> apply(A,1,function(x) sum(x*x))
[1] 107 140 179 224
```

oder man kann die Komponenten der Spalten der Größe nach sortieren:

```
> B=matrix(rnorm(12),4,3)
> B
      [,1]      [,2]      [,3]
[1,] 0.3630699 0.758481970 -0.4143960
[2,] 0.7546151 0.997723025 1.2285883
[3,] -0.3812453 0.004898235 1.2107240
[4,] 0.1462320 -0.615993300 0.0342327
> apply(B,2,sort)
      [,1]      [,2]      [,3]
[1,] -0.3812453 -0.615993300 -0.4143960
[2,] 0.1462320 0.004898235 0.0342327
[3,] 0.3630699 0.758481970 1.2107240
[4,] 0.7546151 0.997723025 1.2285883
```

Die Funktion **sweep** ermöglicht es, auf Zeilen oder Spalten Rechenoperationen anzuwenden. Im folgenden Beispiel dividieren wir die Zeilen einer Matrix durch vorgegebene Werte:

```
> A
      [,1] [,2] [,3]
[1,] 1 5 9
[2,] 2 6 10
[3,] 3 7 11
[4,] 4 8 12
> sweep(A,1,c(1,2,3,4),"/")
      [,1]      [,2]      [,3]
[1,] 1 5.000000 9.000000
[2,] 1 3.000000 5.000000
[3,] 1 2.333333 3.666667
[4,] 1 2.000000 3.000000
```

## Aufgaben

---

(3.1) **Aufgabe:** Es sei

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}.$$

Bestimme  $a_{13}$ ,  $a_{31}$ ,  $a_{33}$ ,  $a_{24}$ . Für welche  $i$  und  $j$  ist  $a_{ij} = 7$ , 10, 4, 12?

**Lösung:**

```
> A=matrix(1:12,3,4)
> which(A==7,arr.ind=TRUE)
      row col
[1,]   1   3
```

(3.2) **Aufgabe:** Bestimme die  $2 \times 3$ -Matrix mit der Eigenschaft  $a_{ij} = i + 2j - 3$ .

**Lösung:**

```
> i=c(1,2)
> j=c(1,2,3)
> A=outer(i,2*j,"+")+3
> A
      [,1] [,2] [,3]
[1,]    6    8   10
[2,]    7    9   11
```

(3.3) **Aufgabe:** Gegeben seien die folgenden Matrizen:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -3 \\ 3 & 4 & 5 \\ 2 & -1 & 7 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 3 & 1 & 5 \\ 1 & -4 & 7 \\ 2 & 1 & 3 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

a) Man bestimme  $\mathbf{D}$  so, daß  $\mathbf{A} + \mathbf{D} = \mathbf{B}$ .

b) Man bestimme  $\mathbf{D}$  so, daß  $\frac{1}{2}(\mathbf{A} - 2\mathbf{D}) + 3\mathbf{B} = \mathbf{D} + 5\mathbf{C}$ .

(3.4) **Aufgabe:** Es sei

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & -3 \\ -1 & 5 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 4 \\ 5 & 1 \end{pmatrix}$$

a) Man berechne  $\mathbf{AB}$ ,  $\mathbf{BA}$ ,  $\mathbf{BC}$ ,  $(\mathbf{AB})\mathbf{C}$ ,  $\mathbf{A}(\mathbf{BC})$ ,  $\mathbf{A}^2$ .

b) Man berechne  $\mathbf{A}^3$ ,  $\mathbf{A}^2 - 5\mathbf{A} + \mathbf{B}$ ,  $\mathbf{A}(\mathbf{B} + \mathbf{C})$ ,  $\mathbf{AB} + \mathbf{AC}$ ,  $(\mathbf{A} + \mathbf{B})^2$ ,  $\mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2$ .

(3.5) **Aufgabe:** Gegeben seien

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

a) Berechne  $\mathbf{A}^2 + \mathbf{AB} + \mathbf{BA} + \mathbf{B}^2$ .

b) Man überprüfe die Distributivgesetze:  $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ ,  $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ . c) Man zeige an einem Beispiel, daß im allgemeinen

$$(\mathbf{A} + \mathbf{B})^2 \neq \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2.$$

(3.6) **Aufgabe:** Überprüfe das Assoziativgesetz der Matrizenmultiplikation anhand von

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & -1 \\ 3 & 0 & 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 3 & 0 & 1 \\ 2 & 1 & 3 \\ 4 & 1 & 2 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \\ -1 & 2 \end{pmatrix}.$$

(3.7) **Aufgabe:** Eine Vermögensanlagegesellschaft bietet dem Anleger an, mit Einzahlungen in jeweils beliebiger Höhe Anteile an drei Portfolios zu erwerben, die ausschließlich aus den vier Standartwerten A, B, C und D zusammengesetzt sind. Im Portfolio  $P_1$  beträgt der

wertmäßige Anteil der Papiere A bis D 70%, 10%, 10% und 10%, in  $P_2$  20%, 30%, 20% und 30% sowie in  $P_3$  10%, 0%, 80% und 10%.

Ein Anleger möchte 250000 GE zu 33%, 18%, 29% und 20% in den Werten A bis D anlegen. Ist dies durch den Erwerb von Anteilen von  $P_1$  bis  $P_3$  möglich? Wenn ja, wieviel muß er in die Portfolios  $P_1$ ,  $P_2$  und  $P_3$  investieren?

**Lösung:** Wir bezeichnen mit  $x_1$ ,  $x_2$  und  $x_3$  die Anteile an  $P_1$ ,  $P_2$  und  $P_3$ , die gekauft werden müssen. Sie sind die Lösung des folgenden linearen Gleichungssystems:

$$\begin{array}{rclcl} 0.7x_1 + 0.2x_2 + 0.1x_3 & = & 82500 & (= 33\% \text{ von } 250000) \\ 0.1x_1 + 0.3x_2 & = & 45000 & (= 18\% \text{ von } 250000) \\ 0.1x_1 + 0.2x_2 + 0.8x_3 & = & 72500 & (= 29\% \text{ von } 250000) \\ 0.1x_1 + 0.3x_2 + 0.1x_3 & = & 50000 & (= 20\% \text{ von } 250000) \\ x_1 + x_2 + x_3 & = & 250000 & \text{Investitionssumme} \end{array}$$

Wir stellen die Gleichungsmatrix auf und wenden das Eliminationsverfahren an:

```
> p1=c(0.7,0.1,0.1,0.1)
> p2=c(0.2,0.3,0.2,0.3)
> p3=c(0.1,0,0.8,0.1)
> p=c(0.33,0.18,0.29,0.2)
> A=rbind(cbind(p1,p2,p3,p*250000),c(1,1,1,250000))
> A
      p1 p2 p3
[1,] 0.7 0.2 0.1 82500
[2,] 0.1 0.3 0.0 45000
[3,] 0.1 0.2 0.8 72500
[4,] 0.1 0.3 0.1 50000
[5,] 1.0 1.0 1.0 250000
> rref(A)$echelon.form
      p1 p2 p3
[1,] 1 0 0 75000
[2,] 0 1 0 125000
[3,] 0 0 1 50000
[4,] 0 0 0 0
[5,] 0 0 0 0
```

Das Gleichungssystem besitzt somit die **eindeutige Lösung**:

$$x_1 = 75000, \quad x_2 = 125000, \quad x_3 = 50000$$

(3.8) **Aufgabe:** Börsenmakler bieten Aktienpakete zum Verkauf an. Der erste bietet ein Paket an, das 20 Aktien der Firma A, 50 der Firma B, 30 der Firma C und 90 der Firma D enthält. In einem Paket des zweiten Händlers befinden sich 100 Aktien der Firma A, 20 der Firma B, 50 der Firma C und 10 der Firma D. Ein Kunde, der seine eigenen Vorstellungen von Risikostreuung hat, möchte gerne 580 Aktien der Firma A, 300 der Firma B, 370 der Firma C und 400 Aktien der Firma D erwerben. Ist es ihm möglich, die gewünschte Zahl von Aktien durch den Ankauf der beiden verschiedenen Sorten Aktienpakete - eventuell von jedem Paket eine gewisse Anzahl - zu bekommen, oder muß er die Aktien einzeln erwerben (was zu einem sehr viel höheren Preis erfolgen müßte)?

(3.9) **Aufgabe:** Eine Vermögensanlagegesellschaft bietet dem Anleger an, mit Einzahlungen in jeweils beliebiger Höhe Anteile an drei Portfolios zu erwerben, die ausschließlich aus

den vier Standardwerten  $A$ ,  $B$ ,  $C$  und  $D$  zusammengesetzt sind. Im Portfolio  $P_1$  beträgt der wertmäßige Anteil der Papiere  $A$  bis  $D$  50%, 0%, 40% und 10%, in  $P_2$  20%, 20%, 50% und 10% sowie in  $P_3$  0%, 30%, 40% und 30%. Ein Anleger möchte ein Portfolio zusammensetzen, in welchem die Werte  $A$ ,  $B$ ,  $C$  und  $D$  mit 22%, 18%, 46% und 14% enthalten sind. Falls dies überhaupt möglich ist, wie muß er die für dieses Investment vorgesehene Summe prozentuell auf  $P_1$ ,  $P_2$  und  $P_3$  verteilen?

## 3.2 Input-Output Analyse

In diesem Abschnitt setzen wir die Einführung in die Input-Output Analyse fort, die wir in der LV Mathematik begonnen haben. Insbesondere wollen wir die Input-Output Analyse in Zukunft mit R durchführen.

### 3.2.1 Die Kostenseite des Input-Output Modells

Wir setzen nun die Untersuchung des Input-Output Modells aus der LV Mathematik fort.

Es sei  $A$  eine Technologiematrix mit der Eigenschaft, daß die Matrix  $E - A$  invertierbar ist und  $(E - A)^{-1}$  nur nichtnegative Elemente besitzt.

Das Input-Output Modell  $x = Ax + b$  beschreibt die Güterseite des Wirtschaftsprozesses. Wenn das Modell auf einem Markt realisierbar sein soll, müssen Preise existieren, die sicherstellen, daß der Markt im Gleichgewicht ist.

Es gibt zwei Arten von Preisen. Erstens gibt es Preise für den Handel mit Gütern. Zweitens gibt es Arbeitslöhne (wages) für die Produktion von Gütern. Es sei

$$w^t = (w_1, w_2, \dots, w_n)$$

der Vektor der Arbeitslöhne, die von den einzelnen Wirtschaftszweigen für die Produktion jeweils einer Mengeneinheit zu bezahlen sind. Außerdem sei

$$p^t = (p_1, p_2, \dots, p_n)$$

der Vektor der Güterpreise, die sowohl für den Handel zwischen den Wirtschaftszweigen als auch für den Endverbrauch gelten sollen.

Wir nehmen an, die Löhne  $w$  seien vorgegeben. Damit der Wirtschaftszweig  $j$  die Kosten der Produktion einer Mengeneinheit decken kann, muß der erzielte Preis für eine produzierte Mengeneinheit mit den dafür anfallenden Kosten übereinstimmen, dh. es muß die Beziehung

$$p_j = a_{1j}p_1 + a_{2j}p_2 + \dots + a_{nj}p_n + w_j \quad (3.1)$$

gelten. Indem man diese Gleichung für alle Wirtschaftszweige fordert und in Matrixschreibweise zusammenfaßt, so erhält man

$$p^t = p^t A + w^t. \quad (3.2)$$

Daraus kann man zum gegebenen Lohnvektor  $w$  den erforderlichen Preisvektor  $p$  durch

$$p^t = w^t (E - A)^{-1} \quad (3.3)$$

berechnen.

Ein Preis-Lohnsystem  $(\mathbf{p}, \mathbf{w})$  ist aber nur dann realisierbar, wenn das Einkommen der Endverbraucher ausreicht, um die für den Endverbrauch bestimmten Gütermengen zu bezahlen. Das Einkommen der Endverbraucher besteht aus den Löhnen für die produzierten Gütermengen  $x_i$  und regelt die Nachfrage:

$$\mathbf{Nachfrage} = w_1x_1 + w_2x_2 + \cdots + w_nx_n = \mathbf{w}^t\mathbf{x}.$$

Diesem Wert der Nachfrage stehen als Angebot die für den Endverbrauch vorgesehenen Gütermengen  $b_i$  mit den entsprechenden Preisen gegenüber:

$$\mathbf{Angebot} = p_1b_1 + p_2b_2 + \cdots + p_nb_n = \mathbf{p}^t\mathbf{b}.$$

Am Markt herrscht Gleichgewicht, wenn Angebot und Nachfrage gleich groß sind. Wir stellen nun die Frage, ob in einem Input-Output Modell diese Gleichgewichtsbedingung erfüllt ist. Der folgende Satz zeigt, daß das immer der Fall ist.

(3.10) SATZ:

Wenn in einem Input-Output Modell die Größen  $\mathbf{x}$ ,  $\mathbf{b}$ ,  $\mathbf{w}$  und  $\mathbf{p}$  durch die Gleichungen

$$\mathbf{x} = \mathbf{Ax} + \mathbf{b} \quad \text{und} \quad \mathbf{p}^t = \mathbf{p}^t\mathbf{A} + \mathbf{w}^t$$

zusammenhängen, dann herrscht Gleichgewicht, dh.  $\mathbf{w}^t\mathbf{x} = \mathbf{p}^t\mathbf{b}$ .

BEGRÜNDUNG: Es gilt immer  $\mathbf{w}^t\mathbf{x} = (\mathbf{p}^t - \mathbf{p}^t\mathbf{A})\mathbf{x} = \mathbf{p}^t(\mathbf{x} - \mathbf{Ax}) = \mathbf{p}^t\mathbf{b}$ . □

(3.11) BEISPIEL:

Gleichgewichtspreise

Wir setzen das Einführungsbeispiel aus dem Mathematik-Skriptum fort. Die Matrix  $(\mathbf{E} - \mathbf{A})^{-1}$  wurde dort berechnet. Wir wollen nun die Kostenseite dieses Modells untersuchen.

Wir nehmen an, daß die Löhne in den drei Wirtschaftssektoren durch den Vektor  $\mathbf{w}^t = (10, 100, 30)$  gegeben sind. Die Produktion einer Tonne Weizen wird also mit 10 GE entlohnt, usw. Damit die Wirtschaft im Gleichgewicht ist, müssen die Preise durch  $\mathbf{p}^t = \mathbf{w}^t(\mathbf{E} - \mathbf{A})^{-1}$ , dh.

$$(10, 100, 30) \begin{pmatrix} 3,4208 & 29,3897 & 11,0814 \\ 0,0922 & 2,2709 & 0,3972 \\ 0,2527 & 3,2976 & 2,2270 \end{pmatrix} = (51,0064, 619,9143, 217,3448)$$

festgelegt werden.

Es sollen nun die Löhne im Sektor Ackerbau um eine Geldeinheit erhöht werden. Wie wirkt sich diese Lohnerhöhung auf die Gleichgewichtspreise aus?

Der Lohnvektor verändert sich um  $\mathbf{e}_1^t = (1, 0, 0)$ . Der neue Preisvektor  $\mathbf{q}^t$  ergibt sich dann aus

$$\begin{aligned} \mathbf{q}^t &= (\mathbf{w}^t + \mathbf{e}_1^t)(\mathbf{E} - \mathbf{A})^{-1} \\ &= \mathbf{w}^t(\mathbf{E} - \mathbf{A})^{-1} + \mathbf{e}_1^t(\mathbf{E} - \mathbf{A})^{-1} \\ &= \mathbf{p}^t + \mathbf{e}_1^t(\mathbf{E} - \mathbf{A})^{-1} \end{aligned}$$

Die Preisänderungen betragen somit

$$\mathbf{q}^t - \mathbf{p}^t = \mathbf{e}_1^t(\mathbf{E} - \mathbf{A})^{-1} = (3,4208, 29,3897, 11,0814)$$

und sind mit der ersten Zeile der Matrix  $(\mathbf{E} - \mathbf{A})^{-1}$  identisch.

## 3.2.2 Ein Musterbeispiel

Als Einstimmung zeigen wir, wie man ein Beispiel aus der Input-Output Analyse mit R lösen kann.

Gegeben ist eine Input-Output Tabelle:

```
> tabelle <- edit(matrix(0,nrow=3,ncol=4))
> tabelle
      col1 col2 col3 col4
[1,]  200   90  120   40
[2,]    7    6    3    5
[3,]   18   12   15   15
```

Wir isolieren die Matrix, die aus den ersten drei Spalten besteht:

```
> T <- tabelle[,1:3]
> T
      col1 col2 col3
[1,]  200   90  120
[2,]    7    6    3
[3,]   18   12   15
```

und den Spaltenvektor des Endbedarfs:

```
> b <- tabelle[,4,drop=FALSE]
> b
      [,1]
[1,]   40
[2,]    5
[3,]   15
```

Dann berechnen wir den aktuellen Outputvektor als Zeilensummen der Input-Output Tabelle:

```
> x <- apply(tabelle,1,sum)
> x
[1] 450  21  60
```

Wir warten noch einen Moment, bevor wir den Outputvektor  $\mathbf{x}$ , der nur ein R-Vektor ist, in einen Spaltenvektor umwandeln. Wir brauchen nämlich den R-Vektor vorher noch zur Berechnung der Technologiematrix.

Um die Technologiematrix zu berechnen, müssen wir die Spalten von  $\mathbf{a}$  durch die Komponenten von  $\mathbf{x}$  dividieren:

```
> A=sweep(T,2,x,"/")
> A
      col1      col2 col3
[1,] 0.44444444 4.2857143 2.00
[2,] 0.01555556 0.2857143 0.05
[3,] 0.04000000 0.5714286 0.25
```

Der Übersichtlichkeit halber stellen wir die Elemente der Technologiematrix  $\mathbf{a}$  als Brüche dar:

```
> library(MASS)
```

```
> fractions(A)
      col1 col2 col3
[1,]  4/9 30/7   2
[2,] 7/450  2/7  1/20
[3,]  1/25  4/7  1/4
```

Nun machen wir auch den Outputvektor  $x$  zu einem Spaltenvektor:

```
> dim(x)=c(3,1)
> x
      [,1]
[1,]  450
[2,]   21
[3,]   60
```

Die Input-Output Gleichung besagt, daß  $x$  mit

```
> A%*%x+b
      [,1]
[1,]  450
[2,]   21
[3,]   60
```

identisch ist.

Nun sollen die Lieferungen des Sektors 2 an den Endverbrauch verdoppelt werden und die Lieferungen der anderen Sektoren halbiert. Wie ist der Output zu planen ?

Der neue Endverbauchsvektor lautet also:

```
> b.neu <- matrix(c(20,10,7.5),3,1)
> b.neu
      [,1]
[1,] 20.0
[2,] 10.0
[3,]  7.5
```

Wir definieren die Einheitsmatrix

```
> E <- diag(c(1,1,1))
> E
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

und berechnen

```
> x.neu <- solve(E-A)%*%b.neu
> x.neu
      [,1]
[1,] 445.42291
[2,]  27.53158
[3,]  54.73233
```

Die Löhne in den drei Wirtschaftssektoren betragen 10, 100 und 30 GE pro Produktion einer Mengeneinheit. Wie müssen die Preise festgelegt werden, damit die Wirtschaft im Gleichgewicht ist ?

Wir bilden den Zeilenvektor der Löhne

```
> w <- matrix(c(10,100,30),nrow=1,ncol=3)
> w
      [,1] [,2] [,3]
[1,]   10  100   30
```

und berechnen

```
> w%%solve(E-A)
      [,1]      [,2]      [,3]
[1,] 51.00642 619.9143 217.3448
```

## 4

## Metrische Konzepte

## 4.1 Skalares Produkt und Norm

**Skalares Produkt**

(4.1) DEFINITION:

Es seien  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ . Unter dem **skalaren** oder **inneren Produkt** von  $\mathbf{x}$  und  $\mathbf{y}$  versteht man die Zahl

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^n x_i y_i \quad \text{wenn } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ und } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Das skalare Produkt von zwei Vektoren ist also eine Zahl, die dadurch entsteht, daß man entsprechende Komponenten der Vektoren multipliziert und die Ergebnisse addiert. Es ist daher klar, daß man nur aus Vektoren der gleichen Dimension ein skalares Produkt bilden kann. Außerdem sollte klar sein, daß man kein skalares Produkt von mehr als zwei Vektoren bilden kann.

Es stellen sich die natürlichen Fragen nach dem Zweck, der Interpretation und den Rechengesetzen des skalaren Produkts. Für die ersten beiden Fragen muß noch ein wenig an die Geduld appelliert werden. Dagegen sind die Rechengesetze einfach zusammenzufassen.

(4.2) SATZ:

- (1) Das skalare Produkt ist symmetrisch, dh.  $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ .
- (2) Das skalare Produkt ist homogen, dh. man kann Multiplikationen mit Zahlen innerhalb oder außerhalb des skalaren Produkts ausführen:  $(\lambda \mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\lambda \mathbf{y}) = \lambda(\mathbf{x} \cdot \mathbf{y})$ .
- (3) Das skalare Produkt ist additiv, dh. man kann die Addition von Vektoren innerhalb oder außerhalb des skalaren Produkts ausführen:

$$(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z} \quad \text{und} \quad \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$$

BEGRÜNDUNG: Die behaupteten Rechengesetze sind so gut wie offensichtlich. Zur Illustration zeigen wir, wie man einen Beweis dafür aufschreibt. Wir beweisen die Eigenschaften (2) und (3) gleichzeitig unter Verwendung des Summenzeichens. Es seien  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  Vektoren mit den Komponenten  $x_i, y_i, z_i$ . Dann gilt für beliebige reelle Zahlen  $\lambda, \mu$

$$\mathbf{x} \cdot (\lambda \mathbf{y} + \mu \mathbf{z}) = \sum_{i=1}^n x_i (\lambda y_i + \mu z_i) = \sum_{i=1}^n (\lambda x_i y_i + \mu x_i z_i)$$

$$= \lambda \sum_{i=1}^n x_i y_i + \mu \sum_{i=1}^n x_i z_i = \lambda(\mathbf{x} \cdot \mathbf{y}) + \mu(\mathbf{x} \cdot \mathbf{z}).$$

□

## Skalarprodukt in R

---

Das Skalarprodukt ist in R sehr einfach zu realisieren:

```
> x=c(1,2,3,4)
> y=c(-1,3,2,0)
> z=c(4,2,-1,-3)
> sum(x*y)
[1] 11
> sum(x*y)
[1] 11
> sum((x+y)*z)
[1] -7
> sum(x*z)+sum(y*z)
[1] -7
```

Man beachte, dass

```
> x*y
[1] -1 6 6 0
```

eine Rechenoperation ist, die einer komponentenweisen Multiplikation ohne Aufsummierung entspricht.

## Norm

---

Wenn man das skalare Produkt eines Vektors mit sich selbst bildet, so entsteht die Größe

$$\mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^n x_i^2 \geq 0.$$

Da diese Größe nichtnegativ ist, kann man daraus die Wurzel ziehen.

(4.3) DEFINITION:

Unter der (**euklidischen**) Norm oder dem **Betrag** eines Vektors  $x \in \mathbb{R}^n$  versteht man

$$\|x\| := \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\sum_{i=1}^n x_i^2}.$$

Die geometrische Interpretation der Norm ist leicht einzusehen.

- Es sei  $\mathbf{x} \in \mathbb{R}^2$  der Vektor einer Verschiebung. Dann ist  $\|\mathbf{x}\|$  die Länge des Verschiebungsvektors.
- Sind  $\mathbf{a} \in \mathbb{R}^2$  und  $\mathbf{b} \in \mathbb{R}^2$  Vektoren, die zu Punkten gehören, dann ist  $\|\mathbf{a} - \mathbf{b}\|$  der Abstand der Punkte.

Um mit der Norm von Vektoren bequem rechnen zu können, benötigen wir die grundlegenden Rechengesetze. Die ersten beiden Rechengesetze sind anschaulich plausibel, wenn man sich die geometrische Interpretation vor Augen hält.

(4.4) SATZ:

- (1)  $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{o}$ .  
 (2)  $\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$ .  
 (3)  $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2$ .

BEGRÜNDUNG: (1) Die erste Aussage ist eine logische Äquivalenz. Ihr Nachweis ist eine gute Übungsaufgabe für formales Beweisen.

(2) Die zweite Aussage kann entweder direkt mit der Definition der Norm bewiesen werden, oder aber man kann auf schon bewiesene Eigenschaften des skalaren Produkts zurückgreifen:

$$\|\lambda\mathbf{x}\|^2 = (\lambda\mathbf{x}) \cdot (\lambda\mathbf{x}) = \lambda^2(\mathbf{x} \cdot \mathbf{x}) = \lambda^2\|\mathbf{x}\|^2.$$

Zur eigentlichen Behauptung gelangt man durch Wurzelziehen. Ist klar, daß  $\sqrt{\lambda^2} = |\lambda|$  ?

(3) Die dritte Aussage muß auf Eigenschaften des skalaren Produkts zurückgeführt werden. Hier wird zum ersten Mal klar, daß das skalare Produkt ein nützliches Konzept ist. Es gilt

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{x} \cdot (\mathbf{x} + \mathbf{y}) + \mathbf{y} \cdot (\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} = \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2. \end{aligned}$$

□

## Norm von Vektoren in R

---

Um die Norm eines Vektors auszurechnen, sind mehrere Rechenoperationen nötig:

```
> x=c(1,2,3,4,5)
> nx=sqrt(sum(x*x))
> nx
[1] 7.416198
```

Um dies zu verkürzen, haben wir in der Library **LinAlg.r** die Funktion **norm** vorgesehen.

```
> norm(x)
[1] 7.416198
```

## Einheitsvektoren und Normierung

---

(4.5) DEFINITION:

Unter einem **Einheitsvektor** versteht man einen Vektor  $\mathbf{x}$  mit der Eigenschaft  $\|\mathbf{x}\| = 1$ .

Ein Einheitsvektor hat also Komponenten, deren Quadratsumme gleich 1 ist. Man kann aus jedem Vektor  $\mathbf{x} \neq \mathbf{o}$  durch Multiplikation mit einer Zahl einen Einheitsvektor machen:

$$\mathbf{y} = \frac{1}{\|\mathbf{x}\|} \mathbf{x} \quad \Longrightarrow \quad \|\mathbf{y}\| = \frac{1}{\|\mathbf{x}\|} \|\mathbf{x}\| = 1$$

Diesen Vorgang nennt man **Normieren**.

```

> x
[1] 1 2 3 4 5
> norm(x)
[1] 7.416198
> y=x/norm(x)
> y
[1] 0.1348400 0.2696799 0.4045199 0.5393599 0.6741999
> norm(y)
[1] 1

```

Typische Einheitsvektoren sind:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

oder

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

## Orthogonale Vektoren

---

Wir kommen nun zum Begriff der Orthogonalität. Die anschauliche Bedeutung dieses Begriffes ist grundlegend für unser geometrisches Vorstellungsvermögen. Orthogonal heißt „senkrecht stehen“ (nicht etwa „lotrecht“ stehen!).

Aus der Abbildung 4.1 entnehmen wir, daß zwei Verschiebungsvektoren  $\mathbf{a}$  und  $\mathbf{b}$  aufeinander senkrecht stehen, wenn

$$\|\mathbf{a} + \mathbf{b}\| = \|\mathbf{a} - \mathbf{b}\|$$

Diese Gleichung hat erstaunliche Konsequenzen. Es gilt nämlich ganz allgemein die Gleichung

$$\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 = 4\mathbf{x} \cdot \mathbf{y}$$

Daraus folgt, dass die Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  genau dann aufeinander senkrecht stehen, wenn  $\mathbf{a} \cdot \mathbf{b} = 0$ .

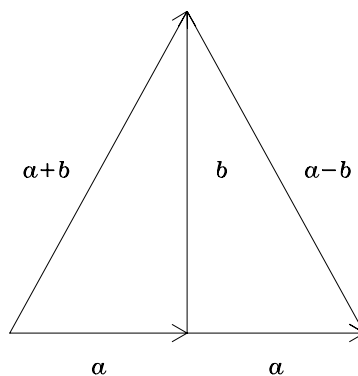


Abbildung 4.1: Orthogonalität

(4.6) DEFINITION:

Zwei Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  heißen **orthogonal zueinander**, wenn  $\mathbf{a} \cdot \mathbf{b} = 0$ .

Sind  $\mathbf{a}$  und  $\mathbf{b}$  orthogonal, so sagt man manchmal,  $\mathbf{a}$  ist ein Normalvektor zu  $\mathbf{b}$ , und  $\mathbf{b}$  ist ein Normalvektor zu  $\mathbf{a}$ . Wir üben den Umgang mit orthogonalen Vektoren anhand einfacher Aufgaben.

## Beispiele in R

---

Offensichtlich orthogonal zueinander sind:

```
> a=c(1,0)
> b=c(0,1)
> sum(a*b)
[1] 0
> coord(xlim=c(0,2),ylim=c(0,2))
> shift(a)
> shift(b)
```

Ein anderes Beispiel:

```
> shift(b)
> a=c(1,1)
> b=c(1,-1)
> sum(a*b)
[1] 0
> coord(xlim=c(0,1),ylim=c(-1,1))
> shift(a)
> shift(b)
```

Wenn die Summe der Komponenten eines Vektors  $\mathbf{x}$  gleich 0 ist, dann bedeutet das, dass der Vektor zu  $\mathbf{e} = (1, 1, \dots, 1)$  orthogonal ist. Das ist zum Beispiel der Fall bei zentrierten Datenvektoren:

```
> x=rnorm(10)
> y=x-mean(x)
> sum(y)
[1] -8.881784e-16
> e=rep(1,10)
> e
[1] 1 1 1 1 1 1 1 1 1 1
> sum(y*e)
[1] -8.881784e-16
```

## Aufgaben

---

(4.7) **Aufgabe:** Gegeben seien die drei Vektoren  $\mathbf{x} = (1, 4, -3, 2, 5, -1)$ ,  $\mathbf{y} = (16, 13, 10, 7, 4, 1)$ ,  $\mathbf{z} = (x_1, x_2, x_3, x_1, x_2, x_3)$  und die Skalare  $\lambda = 0.75$ ,  $\mu = 1.5$ . Berechnen sie die Größe  $(\lambda\mathbf{x} + \mu\mathbf{y})\mathbf{z}$ . Dies ist auf verschiedene Arten möglich.

(4.8) **Aufgabe:** Gegeben seien die zu den Punkten  $A$  und  $B$  gehörenden Vektoren  $\mathbf{a} = (-3, 1)$  und  $\mathbf{b} = (2, -3)$ . Berechnen sie die Länge des Verschiebungsvektors von  $A$  nach  $B$ .

(4.9) **Aufgabe:** Geben sie die normierten Vektoren zu  $\mathbf{x}$ ,  $\mathbf{y}$  und  $\mathbf{z}$  aus Beispiel 4.7. an.

(4.10) **Aufgabe:** Formen Sie um:  $\|\mathbf{x} + \mathbf{y}\|^2$

(4.11) **Aufgabe:** Begründen Sie:  $\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 = 4\mathbf{x} \cdot \mathbf{y}$

(4.12) **Aufgabe:** Es sei  $\mathbf{a} = (1, 3, 2)$  und  $\mathbf{b} = (-1, 5, \lambda)$ . Wählen Sie  $\lambda$  so, dass die beiden Vektoren orthogonal sind.

(4.13) **Aufgabe:** Bestimmen Sie alle zu  $(1, 1)$  orthogonalen Vektoren.

(4.14) **Aufgabe:** Bestimmen Sie alle zu  $(1, -1)$  orthogonalen Vektoren.

(4.15) **Aufgabe:** Beweisen Sie den Satz von Pythagoras: Sind  $\mathbf{a}$  und  $\mathbf{b}$  orthogonale Vektoren im  $\mathbb{R}^n$ , dann gilt  $\|\mathbf{a} + \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$ . (Was hat das mit dem Satz von Pythagoras aus der Schule zu tun?).

(4.16) **Aufgabe:** Ein Deltoid ist ein Viereck, in dem zwei Paare von benachbarten Seiten gleich lang sind. Zeigen Sie, daß die Diagonalen eines Deltoids aufeinander senkrecht stehen.

(4.17) **Aufgabe:** Ein Rhombus ist ein Parallelogramm, bei dem alle Seiten gleich lang sind. Zeigen Sie, daß die Diagonalen eines Rhombus aufeinander senkrecht stehen.

#### (4.18) TESTFRAGEN

1. Was versteht man unter dem Skalarprodukt von Vektoren? Welche Rechengesetze gelten für das Skalarprodukt?
2. Erklären Sie die Begriffe Norm eines Vektors und Abstand zwischen zwei Vektoren? Worin besteht der Zusammenhang? Welche geometrischen Konzepte stehen dahinter?
3. Was ist ein Einheitsvektor? Wie normiert man Vektoren?
4. Erklären Sie den Begriff orthogonal und seine geometrische Motivation.

## 4.2 Orthogonale Projektion

Wir kommen nun zu einem der wichtigsten Begriffe der Mathematik. Es handelt sich dabei um ein Konzept, welches einem Großteil der Methoden der angewandten Mathematik zugrunde liegt.

### Das LSQ-Problem

Es geht um die folgende Problemstellung:

Es seien  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  Vektoren, die beide nicht gleich  $\mathbf{o}$  sind. Wenn die beiden Vektoren nicht proportional sind, dann ist es nicht möglich  $\mathbf{a}$  als Vielfaches von  $\mathbf{b}$  darzustellen. Wenn man  $\mathbf{a}$  mit einem Vielfachen  $\lambda\mathbf{b}$  von  $\mathbf{b}$  vergleicht, bleibt ein Rest, der von  $\mathbf{o}$  verschieden ist:

$$\mathbf{a} = \lambda\mathbf{b} + \mathbf{r} \quad \text{wobei} \quad \mathbf{r} = \mathbf{a} - \lambda\mathbf{b} \neq \mathbf{o}.$$

Für jede Zahl  $\lambda$  erhält man eine andere Darstellung des Vektors  $\mathbf{a}$ .

Wir wollen eine solche Darstellung finden, bei der der Rest  $\mathbf{r}$  möglichst klein ist. Wir wollen also den Vektor  $\mathbf{a}$  möglichst gut durch ein Vielfaches des Vektors  $\mathbf{b}$  approximieren.

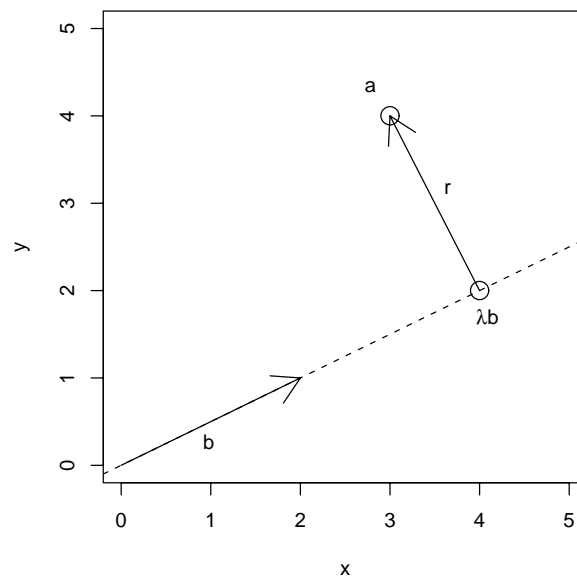


Abbildung 4.2: Orthogonale Projektion

Geometrisch bedeutet das: Bezeichnen wir mit  $[\mathbf{b}]$  die vom Vektor  $\mathbf{b}$  erzeugte Gerade, dh.

$$[\mathbf{b}] = \{\lambda \mathbf{b} : \lambda \in \mathbb{R}\},$$

dann suchen wir jenen Punkt auf dieser Geraden, die zum Punkt  $\mathbf{a}$  den kürzesten Abstand hat. Da der Abstand

$$\|\mathbf{a} - \lambda \mathbf{b}\|^2 = \sum_{i=1}^n (a_i - \lambda b_i)^2$$

die Form einer Quadratsumme hat, nennt man dieses Problem auch das **Least-Squares Problem** (LSQ).

Die Lösung dieses LSQ-Problems ist einfach zu finden.

Zunächst können wir mit R eine numerische Lösung suchen. Wir beziehen uns auf die Situation von Abbildung 4.2.

```
> a=c(3,4)
> b=c(2,1)
> fun=function(t){norm(a-t*b)}
> optimize(fun,c(-5,5))
$minimum
[1] 2.00001
```

```
$objective
[1] 2.236068
```

Offenbar ist  $t = 2$  der optimale Parameter. Die theoretische Antwort ist sehr übersichtlich.

(4.19) SATZ:

**(LSQ-Problem)** Es gibt einen eindeutig bestimmten Punkt  $p_b(\mathbf{a})$  auf der Geraden  $[\mathbf{b}]$ , von dem aus der Abstand zu  $\mathbf{a}$  minimal ist.

Dieser Punkt  $p_b(\mathbf{a})$  heißt die **orthogonale Projektion** von  $\mathbf{a}$  auf  $\mathbf{b}$  und sie hat die Form

$$p_b(\mathbf{a}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b} \quad (4.1)$$

BEGRÜNDUNG: Der gesuchte Punkt  $p_b(\mathbf{a})$  hat die Form  $\lambda \mathbf{b}$ . Wir wollen jene Zahl  $\lambda$  finden, sodass

$$\|\mathbf{a} - \lambda \mathbf{b}\| = \text{Min} !$$

Zu diesem Zweck müssen wir lediglich den Scheitel der quadratischen Funktion

$$f(\lambda) = \|\mathbf{a} - \lambda \mathbf{b}\|^2 = \|\mathbf{a}\|^2 - 2\lambda \mathbf{a} \cdot \mathbf{b} + \lambda^2 \|\mathbf{b}\|^2$$

finden. Die Position des Scheitels beträgt

$$\lambda = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2}.$$

Daher ist

$$p_b(\mathbf{a}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b}.$$

□

Die Formel für die orthogonale Projektion müssen wir uns unbedingt merken. Wir werden sie immer wieder einsetzen, und sie tritt in vielen wichtigen Anwendungen unter anderen Namen auf.

In unserem Beispiel mit  $\mathbf{a} = (3, 4)$  und  $\mathbf{b} = (2, 1)$  ist

$$\lambda_{\text{opt}} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} = \frac{10}{5} = 2.$$

Der Restvektor lautet

$$\mathbf{r} = \mathbf{a} - \lambda_{\text{opt}} \mathbf{b} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - 2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Bemerkenswert daran ist, dass

$$\mathbf{r} \cdot \mathbf{b} = 0$$

ist, dh. dass  $\mathbf{r}$  und  $\mathbf{b}$  orthogonal sind. Dahinter steht ein allgemeines Gesetz.

## Orthogonale Zerlegung

---

In der Abbildung 4.2 sehen wir, dass der Verbindungsvektor

$$\mathbf{r} = \mathbf{b} - p_b(\mathbf{a})$$

zu  $\mathbf{b}$  und daher zur gesamten Geraden  $[\mathbf{b}]$  orthogonal ist.

Wie wir sehen werden, hat die optimale Zerlegung  $\mathbf{a} = p_b(\mathbf{a}) + \mathbf{r}$  aber noch eine weitere Besonderheit: Es ist die einzige Zerlegung, bei der der Rest  $\mathbf{r}$  zum Vektor  $\mathbf{b}$  orthogonal ist.

Dieser Sachverhalt ist von großer Bedeutung. Wir fassen ihn daher in einem weiteren Theorem zusammen.

(4.20) SATZ:

**(Orthogonale Zerlegung)**

(1) Ist  $\mathbf{b} \neq \mathbf{o}$ , dann läßt sich jeder Vektor  $\mathbf{a}$  als Summe  $\mathbf{a} = \mathbf{u} + \mathbf{v}$  zweier Vektoren  $\mathbf{u}, \mathbf{v}$  darstellen, von denen  $\mathbf{u}$  zu  $\mathbf{b}$  proportional ist und  $\mathbf{v}$  zu  $\mathbf{b}$  orthogonal ist. Man nennt eine solche Darstellung eine **orthogonale Zerlegung**.

(2) Die Zerlegung  $\mathbf{a} = p_b(\mathbf{a}) + \mathbf{r}$  ist eine solche orthogonale Zerlegung, und sie ist die einzig mögliche.

BEGRÜNDUNG: Es sei  $\mathbf{a} = \lambda \mathbf{b} + \mathbf{r}$ . Dann gilt

$$\lambda = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \Leftrightarrow \lambda \mathbf{b} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b} \Leftrightarrow (\mathbf{a} - \lambda \mathbf{b}) \cdot \mathbf{b} = \mathbf{o} \Leftrightarrow \mathbf{r} \cdot \mathbf{b} = \mathbf{o}$$

Die Bedingung  $\mathbf{r} \cdot \mathbf{b} = \mathbf{o}$  ist also äquivalent mit  $\lambda \mathbf{b} = p_b(\mathbf{a})$ . □

Für orthogonale Zerlegungen gilt der **Satz von Pythagoras**:

(4.21) SATZ:

Ist  $\mathbf{a} = \mathbf{u} + \mathbf{v}$  eine orthogonale Zerlegung, dann gilt

$$\|\mathbf{a}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \quad (4.2)$$

BEGRÜNDUNG: Durch Nachrechnen. □

Wir werden bald sehen, dass in vielen Anwendungen orthogonale Zerlegungen vorkommen. Hochmoderne Konzepte der Mathematik haben so gesehen den gleichen Grundgedanken wie der Satz von Pythagoras, der bereits den babylonischen Landvermessern bekannt war.

## Der Winkel zwischen Vektoren

Es ist anschaulich klar, dass eine orthogonale Projektion  $p_b(\mathbf{a})$  umso besser ist (dh. dass der Rest  $\mathbf{r}$  umso kleiner ist), je kleiner der Winkel zwischen den Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  ist. Wir wollen diesen Gedanken nun quantitativ untersuchen.

Es ist eine uralte Idee in der Geometrie, Winkel dadurch zu messen, dass man orthogonale Projektion misst. Das Ergebnis sind die aus der Schule bekannten Begriffe wie **sinus** und **cosinus**. Das soll nun näher erklärt werden.

Gegeben sind zwei Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ . Durch einfaches Nachrechnen sehen wir: Das Verhältnis zwischen der Länge der Projektion  $p_b(\mathbf{a})$  und der Länge des projizierten Vektors beträgt

$$\frac{\|p_b(\mathbf{a})\|}{\|\mathbf{a}\|} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} =: c(\mathbf{a}, \mathbf{b})$$

Die dabei auftretende Größe

$$c(\mathbf{a}, \mathbf{b}) := \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \quad (4.3)$$

ist geometrisch gesehen nicht anderes als der **cosinus** des Winkels zwischen den Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ .

Es sei  $\mathbf{a} = p_b(\mathbf{a}) + \mathbf{r}$ . Dann folgt aus dem Satz von Pythagoras:

$$\|\mathbf{a}\|^2 = \|p_b(\mathbf{a})\|^2 + \|\mathbf{r}\|^2 \quad \Rightarrow \quad \|\mathbf{r}\| \leq \|\mathbf{a}\|$$

Der Rest  $\mathbf{r}$  ist also immer kleiner (oder wenigstens nicht größer) als der Vektor  $\mathbf{a}$ , der projiziert wird. Das Ausmaß, in dem  $\|\mathbf{r}\|$  kleiner ist als  $\|\mathbf{a}\|$ , läßt sich genau quantifizieren. Mit dem Satz von Pythagoras erhält man nämlich eine Formel, die man in der Schule den **Cosinussatz** nennt:

$$\begin{aligned}\|\mathbf{r}\|^2 &= \|\mathbf{a}\|^2 - \|p_{\mathbf{b}}(\mathbf{a})\|^2 \\ &= \|\mathbf{a}\|^2 \left(1 - \frac{\|p_{\mathbf{b}}(\mathbf{a})\|^2}{\|\mathbf{a}\|^2}\right) \\ &= \|\mathbf{a}\|^2 (1 - c(\mathbf{a}, \mathbf{b})^2)\end{aligned}\quad (4.4)$$

Die Gleichung ist die Grundlage für eine Vielzahl späterer statistischer Anwendungen.

Zunächst aber folgt daraus die berühmte **Ungleichung von Cauchy–Schwarz**. Diese Ungleichung enthält Information darüber, welche Werte ein skalares Produkt im Extremfall annehmen kann.

(4.22) SATZ:

(Ungleichung von Cauchy–Schwarz) Es seien  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ . Dann gilt

$$|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|.$$

BEGRÜNDUNG: In der Gleichung 4.4 ist die linke Seite immer  $\geq 0$ . Daher muss auf der rechten Seite  $c(\mathbf{a}, \mathbf{b})^2 \leq 1$  gelten.  $\square$

Die Ungleichung von Cauchy–Schwarz wurde in früheren Zeiten natürlich nicht so abstrakt formuliert. Sie lautet in ausführlicher Komponentenschreibweise

$$\left| \sum_{i=1}^n a_i b_i \right| \leq \sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}.$$

In dieser Form ist die Ungleichung seit dem 19. Jahrhundert bekannt.

## Aufgaben

(4.23) **Aufgabe:** Was besagt die Ungleichung von Cauchy–Schwarz im Fall  $n = 1$ ?

(4.24) **Aufgabe:** Zeigen Sie:  $|\mathbf{a} \cdot \mathbf{b}| = \|\mathbf{a}\| \|\mathbf{b}\| \Leftrightarrow \mathbf{a} \parallel \mathbf{b}$ .

Hinweis: Untersuchen Sie, wann in der orthogonalen Zerlegung  $\mathbf{a} = \lambda \mathbf{b} + \mathbf{r}$  der Fall  $\|\mathbf{r}\| = 0$  eintritt.

(4.25) **Aufgabe:** Der Betrag des skalaren Produkts kann positiv oder negativ sein. Daher kann das Gleichheitszeichen in der Ungleichung von Cauchy–Schwarz auf zwei Arten zustandekommen. Beide Fälle sind geometrisch charakterisierbar. Zeigen Sie:

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \Leftrightarrow \mathbf{a} \text{ und } \mathbf{b} \text{ sind gleich gerichtet}$$

$$\mathbf{a} \cdot \mathbf{b} = -\|\mathbf{a}\| \|\mathbf{b}\| \Leftrightarrow \mathbf{a} \text{ und } \mathbf{b} \text{ sind entgegengesetzt gerichtet}$$

(4.26) **Aufgabe:** Es seien  $\mathbf{a} \neq \mathbf{o}$  und  $\mathbf{b} \neq \mathbf{o}$  Vektoren in  $\mathbb{R}^n$ . Zeigen Sie:

(1)  $c(\mathbf{a}, \mathbf{b})$  liegt immer zwischen  $-1$  und  $1$ .

(2)  $c(\mathbf{a}, \mathbf{b}) = 0$  genau dann, wenn  $\mathbf{a}$  und  $\mathbf{b}$  orthogonal sind.

(3)  $c(\mathbf{a}, \mathbf{b}) = 1$  genau dann, wenn  $\mathbf{a}$  und  $\mathbf{b}$  gleich gerichtet sind.

(4)  $c(\mathbf{a}, \mathbf{b}) = -1$  genau dann, wenn  $\mathbf{a}$  und  $\mathbf{b}$  entgegengesetzt gerichtet sind.

(4.27) **Aufgabe:** Beweisen Sie die Ungleichung von Minkowski:

$$\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|.$$

Hinweis: Rechnen Sie  $\|\mathbf{a} + \mathbf{b}\|^2$  aus und wenden Sie dann die Ungleichung von Cauchy–Schwarz an.

(4.28) **Aufgabe:** Interpretieren Sie die Ungleichung von Minkowski als Dreiecksungleichung.

(4.29) **Aufgabe:** Analysieren Sie, unter welchen Umständen in der Ungleichung von Minkowski das Gleichheitszeichen gilt.

(4.30) **Aufgabe:** Was sagen Sie zu den folgenden Ungleichungen ?

(1)  $\|\mathbf{a} - \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$

(2)  $\|\mathbf{a} - \mathbf{b}\| \leq \|\mathbf{a}\| - \|\mathbf{b}\|$

(3)  $\|\mathbf{a}\| - \|\mathbf{b}\| \leq \|\mathbf{a} - \mathbf{b}\|$

#### (4.31) TESTFRAGEN

1. Was ist eine orthogonale Projektion eines Vektors auf einen anderen Vektor ? Erklären sie das Problem und seine Lösung in geometrischer Sichtweise.
2. Was hat das LSQ-Prinzip mit einer orthogonalen Projektion zu tun ?
3. Wie berechnet man eine orthogonale Projektion eines Vektors auf einen anderen Vektor ?
4. Was ist eine orthogonale Zerlegung, und was haben orthogonale Zerlegungen mit Projektionen zu tun ?
5. Wie kann man das Skalarprodukt mit Hilfe des Begriffs einer orthogonalen Projektion anschaulich deuten ?
6. Was versteht man unter dem Cosinus des Winkels zwischen zwei Vektoren ? Geben Sie zwei Deutungen.
7. Welche Eigenschaften hat der Cosinus des Winkels zwischen zwei Vektoren ?
8. Erläutern Sie das Problem der kürzesten Verbindung eines Punktes zu einer Geraden.
9. Begründen Sie, warum die kürzeste Verbindungsstrecke eines Punktes zu einer Geraden stets zur Geraden orthogonal sein muß ?
10. Wie lautet die Ungleichung von Cauchy–Schwarz ? Wann gilt dort das Gleichheitszeichen ? Wie nützt man diese Ungleichung für die geometrische Interpretation des skalaren Produkts ?
11. Wie lautet die Ungleichung von Minkowski ? Wann gilt dort das Gleichheitszeichen ? Welche geometrische Interpretation besitzt die Ungleichung von Minkowski ?

## 4.3 Statistische Anwendungen

### 4.3.1 Der Mittelwert

Es sei  $\mathbf{x} \in \mathbb{R}^n$  eine Datenliste. Wir wollen den Mittelwert in Vektorschreibweise darstellen. Zu diesem Zweck definieren wir den Vektor

$$\mathbf{e} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Mit diesem Vektor kann man eine Summe als Skalarprodukt schreiben:

$$\sum_{i=1}^n x_i = \mathbf{x} \cdot \mathbf{e}.$$

Der Vektor  $\mathbf{e}$  hat die Norm  $\|\mathbf{e}\| = \sqrt{n}$ . Daher ist

$$\bar{x} = \frac{\mathbf{x} \cdot \mathbf{e}}{\|\mathbf{e}\|^2}. \quad (4.5)$$

Daraus erkennen wir den folgenden Zusammenhang:

**Ein Datenvektor  $\mathbf{x}$  hat genau dann den Mittelwert  $\bar{x} = 0$  (dh, ist zentriert), wenn er zum Vektor  $\mathbf{e}$  orthogonal ist.**

Darüber hinaus hat die auf den ersten Blick umständliche Schreibweise 4.5 eine tiefere Bedeutung. Man erkennt daraus, dass  $\bar{x}$  mit einer orthogonalen Projektion zusammenhängt:

$$p_{\mathbf{e}}(\mathbf{x}) = \frac{\mathbf{x} \cdot \mathbf{e}}{\|\mathbf{e}\|^2} \mathbf{e} = \bar{x} \mathbf{e}.$$

Daher ist die Gleichung

$$\mathbf{x} = \bar{x} \mathbf{e} + \mathbf{r}$$

eine orthogonale Zerlegung, und der Vektor der sogenannten **Residuen**

$$\mathbf{r} = \mathbf{x} - \bar{x} \mathbf{e} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$$

ist **orthogonal zu  $\mathbf{e}$** , dh. er ist **zentriert**.

### Hinweise zu R

---

Mittelwerte berechnet man durch

```
> mean(MBAG)
[1] 28.27778
```

Residuen entstehen durch

```
> y=MBAG-mean(MBAG)
> mean(y)
[1] -4.200207e-14
> round(mean(y),5)
[1] 0
```

Manchmal will man die Mittelwerte aller Spalten einer Matrix berechnen:

```
> xx=cbind(MBAG,FBAG,FIB,FIT)
> apply(xx,2,mean)
      MBAG      FBAG      FIB      FIT
28.27778 31.51080 73.83565 155.19444
```

Auch das Zentrieren ist spaltenweise möglich:

```
yy=scale(xx,center=TRUE,scale=FALSE)
```

Allerdings sind diese voreingestellten Funktionen nicht sehr flexibel. Viel leistungsfähiger ist die Funktion **sweep**.

```
> yy=sweep(xx,2,apply(xx,2,mean),"-")
> yy[1:10,]
      MBAG      FBAG      FIB      FIT
[1,] -11.27778 -12.510802 -40.83565  -5.194444
[2,] -11.27778  -8.510802 -33.83565  19.805556
[3,] -10.27778 -10.510802 -29.83565 -39.194444
[4,] -10.27778  -5.510802 -31.83565 -43.194444
[5,] -10.27778 -10.510802 -23.83565 -26.194444
[6,] -10.27778 -14.510802 -73.83565  58.805556
[7,] -10.27778  -8.510802 -18.83565 -13.194444
[8,] -10.27778 -11.510802 -35.83565 -35.194444
[9,] -15.27778  -8.510802 -39.83565 -51.194444
[10,] -10.27778  -3.510802 -25.83565  38.805556
```

```
> yy=sweep(xx,2,10,"-")
> yy[1:10,]
      MBAG FBAG FIB FIT
[1,]    7    9  23 140
[2,]    7   13  30 165
[3,]    8   11  34 106
[4,]    8   16  32 102
[5,]    8   11  40 119
[6,]    8    7 -10 204
[7,]    8   13  45 132
[8,]    8   10  28 110
[9,]    3   13  24  94
[10,]   8   18  38 184
```

## 4.3.2 Die Varianz

Die Varianz eines Datenvektors  $\mathbf{x}$  hängt eng mit der Norm (der Länge) des zentrierten Datenvektors  $\mathbf{x} - \bar{x}\mathbf{e}$  zusammen.

Zunächst ist klar, dass

$$\|\mathbf{x} - \bar{x}\mathbf{e}\|^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

Die Varianz  $s_x^2$  entsteht dann dadurch, dass noch durch einen Faktor dividiert wird. Dieser Faktor ist im Fall der Varianz nicht ganz eindeutig.

Wir verwenden im folgenden die Stichprobenvarianz, dh.

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\|\mathbf{x} - \bar{x}\mathbf{e}\|^2}{n-1}$$

## Hinweise zu R

Varianz und Standardabweichung bildet man durch

```
> var(MBAG)
[1] 36.03089
> sd(MBAG)
[1] 6.002573
```

Mit **apply** kann man das auch spaltenweise machen. Standardisierung erfolgt mit **scale**.

### 4.3.3 Kovarianz und Korrelation

Wir wenden uns nun der Deutung des Korrelationskoeffizienten zu.

Eine Deutung des Korrelationskoeffizienten haben wir schon in der Statistik LV kennengelernt. Die Deutung beruht auf dem Streudiagramm, also der Darstellung der **Zeilen** der Datenmatrix als Punkte.

Wir werden nun einer andere Deutung des Korrelationskoeffizienten kennenlernen, die darauf beruht, daß wir die **Spalten** der Datenmatrix als Vektoren auffassen.

(4.32) SATZ:

*Der Korrelationskoeffizient von zwei Datenvektoren  $\mathbf{x}$  und  $\mathbf{y}$  ist der Cosinus des Winkels zwischen den zentrierten Datenvektoren  $\mathbf{x} - \bar{x}\mathbf{e}$  und  $\mathbf{y} - \bar{y}\mathbf{e}$ , dh.*

$$r_{xy} = c(\mathbf{x} - \bar{x}\mathbf{e}, \mathbf{y} - \bar{y}\mathbf{e}) = \frac{(\mathbf{x} - \bar{x}\mathbf{e}) \cdot (\mathbf{y} - \bar{y}\mathbf{e})}{\|\mathbf{x} - \bar{x}\mathbf{e}\| \|\mathbf{y} - \bar{y}\mathbf{e}\|} \quad (4.6)$$

BEGRÜNDUNG: Wir formen einfach die uns vertraute Definition um:

$$r_{xy} = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{n}s_x \sqrt{n}s_y} = \frac{(\mathbf{x} - \bar{x}\mathbf{e}) \cdot (\mathbf{y} - \bar{y}\mathbf{e})}{\|\mathbf{x} - \bar{x}\mathbf{e}\| \|\mathbf{y} - \bar{y}\mathbf{e}\|}$$

□

Aus dieser Darstellung des Korrelationskoeffizienten ergibt sich für uns die folgende neue Interpretation:

**Der Korrelationskoeffizient  $r_{xy}$  mißt die Ähnlichkeit zwischen den Richtungen der Datenvektoren  $\mathbf{x}$  und  $\mathbf{y}$ , dh. zwischen den Datenvektoren unabhängig von ihrer Länge (ihrer Streuung).**

Für statistische Anwendungen dividiert man Zähler und Nenner in 4.6 durch  $n - 1$ , wie wir das schon bei der Varianz gemacht haben. Wir erhalten so die **Kovarianz** zwischen  $x$  und  $y$ :

$$s_{xy} = \frac{(\mathbf{x} - \bar{x}\mathbf{e}) \cdot (\mathbf{y} - \bar{y}\mathbf{e})}{n - 1} = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Damit erhält der Korrelationskoeffizient die Form

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

## Hinweise zu R

---

Kovarianz und Korrelation sind einfach zu berechnen:

```
> cov(MBAG, FBAG)
[1] 33.06109
> cor(MBAG, FBAG)
[1] 0.8268894
```

Wichtiger ist die Berechnung der Kovarianzmatrix und der Korrelationsmatrix einer Datenmatrix:

```
> cov(xx)
      MBAG      FBAG      FIB      FIT
MBAG 36.03089  33.06109  49.86885  20.59151
FBAG 33.06109  44.36745  52.56664 -14.69631
FIB  49.86885  52.56664 861.99382  648.01345
FIT  20.59151 -14.69631 648.01345 4657.19768
> cor(xx)
      MBAG      FBAG      FIB      FIT
MBAG 1.0000000  0.8268894 0.2829696  0.05026762
FBAG 0.8268894  1.0000000 0.2687979 -0.03233060
FIB  0.2829696  0.2687979 1.0000000  0.32342204
FIT  0.05026762 -0.0323306 0.3234220  1.00000000
```

## Aufgaben

---

### (4.33) TESTFRAGEN

1. Welche geometrische Interpretation besitzt das skalare Produkt eines Vektors mit einem Einheitsvektor ?
2. Wie lautet die vektorielle Darstellung eines Mittelwerts ?
3. Wie lautet die vektorielle Darstellung der Varianz ?
4. Wie lautet die vektorielle Darstellung der Varianz des Korrelationskoeffizienten ?

## 5

## Die Methode der kleinsten Quadrate

## 5.1 Das einfache Regressionsmodell

## 5.1.1 Modellanpassung (model fitting)

Beim einfachen Regressionsmodell geht es darum, eine Responsevariable  $Y$  durch eine Prädiktorvariable  $X$  zu erklären, dh. möglichst gut vorherzusagen.

Es liege eine Stichprobe vom Umfang  $n$  vor, also Daten  $y_1, y_2, \dots, y_n$  der Responsevariablen und Daten  $x_1, x_2, \dots, x_n$  der Prädiktorvariablen.

(5.1) DEFINITION:

Unter einem **einfachen linearen Regressionsmodell** versteht man ein Modell der Form

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Dabei bezeichnen  $\beta_0$  und  $\beta_1$  die **Modellparameter** und  $\epsilon_i$  die **Residuen**.

Mit einem solchen Regressionsmodell versucht man dann, die Daten  $y_i$  durch die lineare Funktion  $\hat{y}_i = \beta_0 + \beta_1 x_i$  vorherzusagen. Der Fehler, den man dabei macht, ist

$$y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x_i = \epsilon_i$$

Das sind gerade die Residuen. Das Problem besteht daher darin, ein optimales Modell zu finden, dh. ein Modell, bei dem die Residuen möglichst klein sind.

Die wichtigste und einfachste Methode, ein optimales Regressionsmodell zu finden, ist die **Methode der kleinsten Quadrate**. Man nennt sie auch das **Least Squares** Prinzip, und kürzt es als **LSQ** ab.

(5.2) DEFINITION:

Unter dem **LSQ-Prinzip** versteht man die Optimierungsaufgabe

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = \text{Min} !$$

Die Modellparameter sollen also so gewählt werden, dass die Quadratsumme der Residuen minimal ist.

Die optimalen Modellparameter  $\hat{\beta}_0$  und  $\hat{\beta}_1$  heißen **Regressionsparameter**. Die optimale Prognosefunktion

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

ist die **lineare Regressionsfunktion**. Auch für die speziellen Residuen, die beim optimalen Modell anfallen, verwenden wir ein besonderes Symbol:

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

Sehen wir uns ein Beispiel an.

### (5.3) BEISPIEL:

Der Responsevektor  $\mathbf{y}$  soll mit einem linearen Regressionsmodell durch den Prädiktorvektor  $\mathbf{x}$  erklärt werden.

```
> x
[1] 1 2 3 4 5 6 7 8 9 10
> y
[1] -4 -3 1 6 10 7 0 8 17 14
```

1. Zeichnen Sie das Streudiagramm.
2. Zeichnen Sie die Prognosefunktion für die Modellparameter  $\beta_0 = 0.1$  und  $\beta_1 = 1$  ein. Wie lautet der Vektor der Residuen ?
3. Bestimmen Sie die Regressionsparameter  $\hat{\beta}_0$  und  $\hat{\beta}_1$ . Zeichnen Sie die Regressionsfunktion ein. Wie lautet der Vektor der Residuen ?
4. Vergleichen Sie die Quadratsummen der Residuen der beiden Modelle.

### Lösung:

Das Streudiagramm erhalten wir durch

```
> plot(x,y)
```

Nun zeichnen wir das Modell  $y = 0.1 + x$  ein:

```
> abline(0.1,1)
```

Es ist sogar sehr einfach die Residuen graphisch darzustellen. Man muss zu diesem Zweck nur die Datenpunkte des Streudiagramms mit den entsprechenden Punkten auf der Modellgeraden verbinden:

```
> arrows(x,y,x,0.1+x,length=0)
```

Die Residuen lauten

```
> res.1=y-(0.1+x)
> res.1
[1] -5.1 -5.1 -2.1 1.9 4.9 0.9 -7.1 -0.1 7.9 3.9
> norm(res.1)^2
[1] 212.9
```

Die optimalen Parameter, also die Regressionsparameter berechnen wir einstweilen noch mit den bekannten Formeln aus der Statistik LV:

```
> beta.1=sd(y)/sd(x)*cor(x,y)
> beta.0=mean(y)-beta.1*mean(x)
> beta.1
```

```
[1] 1.915152
> beta.0
[1] -4.933333
```

Die Regressionsgerade und ihre Residuen zeichnen wir nun in einer anderen Farbe ein:

```
> abline(beta.0,beta.1,col="red")
> arrows(x,y,x,beta.0+beta.1*x,length=0,col="red")
```

Schließlich noch die Residuen:

```
> res.2=y-(beta.0+beta.1*x)
> norm(res.2)^2
[1] 143.8061
```

## LSQ mit R

Für die Lösung und Auswertung von LSQ-Problemen stellt R ein sehr leistungsfähiges Instrument zu Verfügung. Wir beginnen jetzt damit, dieses Instrument näher kennenzulernen.

Eine lineares Regressionsmodell wird in R durch eine Modellformel definiert: Man benennt die Responsevariable und trennt mit „~“ die darauffolgende Benennung der Prädiktervariablen:  $y \sim x$ .

Die Berechnung des optimalen Modells erfolgt dann durch die Funktion **lm** (linear model). Diese Funktion berechnet außerdem einige wichtige Zwischenergebnisse. Daher ist es am besten, wenn man das Ergebnis der Funktion in einer Variablen (z.B. **fm**, fitted model) abspeichert, um beim Aufruf der Ausgabefunktionen das Modell nicht immer wieder neu berechnen zu müssen.

Wenn wir das mit den Zahlen unseres Beispiels 5.3 tun, dann erhalten wir:

```
> coef(fm)
(Intercept)          x
  -4.933333    1.915152
> residuals(fm)
      1          2          3          4          5
-0.9818182 -1.8969697  0.1878788  3.2727273  5.3575758
      6          7          8          9         10
  0.4424242 -8.4727273 -2.3878788  4.6969697 -0.2181818
> fitted(fm)
      1          2          3          4          5
-3.0181818 -1.1030303  0.8121212  2.7272727  4.6424242
      6          7          8          9         10
  6.5575758  8.4727273 10.3878788 12.3030303 14.2181818
```

### (5.4) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: FBAG, Prädiktor: MBAG.

Welches Alter des Vaters ist bei einer 33-jährigen Mutter im Durchschnitt zu erwarten ?

**Lösung:**

Wir berechnen das Modell:

```
> fm=lm(FBAG~MBAG)
```

Dadurch erhalten wir die Regressionskoeffizienten:

```
> coef(fm)
(Intercept)          MBAG
  5.5637832      0.9175763
```

Wir können jetzt das Streudiagramm zeichnen und die Regressionsgerade einzeichnen:

```
> plot(MBAG, FBAG)
> abline(coef(fm)[1], coef(fm)[2])
```

Die Prognose lautet:

```
> coef(fm)[1]+33*coef(fm)[2]
(Intercept)
  35.8438
```

## 5.1.2 Die mathematische Lösung

Wir sehen uns jetzt an, wie man zu den Lösungsformeln für die Regressionsparameter kommt.

### Das Absolutglied

Bei Regressionsmodellen spielt das Absolutglied  $\beta_0$  des Modells immer eine Sonderrolle. Aus dem LSQ-Prinzip kann man nämlich leicht ablesen, dass  $\hat{\beta}_0$  aus den anderen Regressionsparametern  $\hat{\beta}_i$  ganz einfach berechnet werden kann.

Wir haben diesen Umstand schon in der Statistik LV kennengelernt. Dort haben wir darauf hingewiesen, dass die Regressionsgerade stets durch den Mittelpunkt der Punktwolke führt. Daher mussten wir uns nur die Formel für  $\hat{\beta}_1$  merken. Nun lernen wir, was hinter dieser Tatsache steckt.

**Bei einer optimalen Lösung des LSQ-Problems muss die Summe der Residuen (und daher auch ihr Mittelwert) stets gleich Null sein !**

Bevor wir von dieser Regel Gebrauch machen, versuchen wir, sie zu verstehen. Angenommen, wir hätten optimale Modellparameter  $\hat{\beta}_0$  und  $\hat{\beta}_1$ , aber die Residuen  $\epsilon_i$ , die dazu gehören, haben einen Mittelwert  $\bar{\epsilon} \neq 0$ . Dann hätte von den beiden Modellen

$$\begin{aligned} y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i \\ &= (\hat{\beta}_0 + \bar{\epsilon}) + \hat{\beta}_1 x_i + (\epsilon_i - \bar{\epsilon}) \end{aligned}$$

das zweite Modell eine kleinere Quadratsumme der Residuen. Das würde aber der unterstellten Optimalität des ersten Modells widersprechen.

Der Vektor der (optimalen) Residuen ist also immer zentriert. Wenn wir in der Modellgleichung

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\epsilon}_i \tag{5.1}$$

auf beiden Seiten den Mittelwert bilden, fallen daher die Residuen weg und wir erhalten

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad (5.2)$$

Diese Gleichung ist die mathematische Fassung der Aussage, dass die Regressionsgerade durch den Mittelpunkt der Punktwolke geht.

Aus der Gleichung 5.2 können wir ablesen, dass der Regressionsparameter  $\hat{\beta}_0$  sofort berechnet werden kann, wenn man den anderen Regressionsparameter  $\hat{\beta}_1$  kennt:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Für die Festlegung des optimalen Modells genügt es daher, nur den zweiten Modellparameter  $\beta_1$  zu optimieren.

Wir subtrahieren die Gleichungen 5.1 und 5.2 und erhalten eine vereinfachte Form der Modellgleichung:

$$y_i - \bar{y} = \hat{\beta}_1(x_i - \bar{x}) + \hat{\epsilon}_i \quad (5.3)$$

In dieser Modellgleichung kommt jetzt nur mehr der zweite Modellparameter  $\beta_1$  vor.

## LSQ Optimierung als orthogonale Projektion

---

Die Modellgleichung 5.3 zeigt uns, dass wir in der Gleichung

$$y_i - \bar{y} = \beta_1(x_i - \bar{x}) + \epsilon_i$$

den Parameter  $\beta_1$  nach dem LSQ Prinzip optimieren müssen. In ausführlicher Schreibweise bedeutet das

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n [(y_i - \bar{y}) - \beta_1(x_i - \bar{x})]^2 = \text{Min} !$$

oder in Vektorschreibweise

$$\|\epsilon\|^2 = \|(\mathbf{y} - \bar{y}\mathbf{e}) - \beta_1(\mathbf{x} - \bar{x}\mathbf{e})\|^2 = \text{Min} !$$

Dieses Optimierungsproblem ist aber nichts anderes als die Grundaufgabe der orthogonalen Projektion. Und daher kennen wir die optimale Lösung für  $\beta_1$ :

$$\hat{\beta}_1 = \frac{(\mathbf{y} - \bar{y}\mathbf{e}) \cdot (\mathbf{x} - \bar{x}\mathbf{e})}{\|\mathbf{x} - \bar{x}\mathbf{e}\|^2} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{s_y}{s_x} r_{yx}$$

Aus dem Umstand, dass es sich bei der optimalen Lösung des LSQ Problems um eine orthogonale Projektion handelt, können wir eine zweite Eigenschaft der optimalen Residuen folgern:

**Bei einer optimalen Lösung des LSQ Problems müssen die Residuen  $\hat{\epsilon}_i = y_i - \hat{y}_i$  zu den Prädiktoren  $x_i$  unkorreliert sein, dh. der Vektor  $\hat{\epsilon}$  muß zu  $\mathbf{x} - \bar{x}\mathbf{e}$  orthogonal sein.**

## Die Normalgleichungen

---

Wir werden später multiple Regressionsmodelle, dh. Regressionsmodelle mit mehr als einem Prädiktor behandeln. Bei solchen komplizierteren Modellen wird die optimale Lösung durch ein lineares Gleichungssystem, die sogenannten **Normalgleichungen**, gewonnen. Als Vorbereitung dafür wollen wir uns schon jetzt die Struktur dieses Gleichungssystems ansehen. (Natürlich ist im Fall der einfachen Regression die Berechnung der orthogonalen Projektion der viel einfachere Weg !)

(5.5) SATZ:

*Das System der Normalgleichungen hat die Form*

$$\mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{y}$$

wobei

$$\mathbf{X} := \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

*die sogenannte **Designmatrix** des linearen Regressionsmodells bezeichnet.*

Bevor wir erklären, wie dieses System von Normalgleichungen entsteht, wollen wir es praktisch erproben.

---

(5.6) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: FBAG, Prädiktor: MBAG.

Berechnen Sie die Regressionsparameter mit dem System der Normalgleichungen.

**Lösung:** Wir berechnen zunächst die Designmatrix:

```
> x=MBAG
> y=FBAG
> e=rep(1,1296)
> X=cbind(e,x)
```

Die Lösung der Normalgleichungen erhalten wir durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$$

also

```
> beta=solve(t(X)**X)**t(X)**y
> beta
      [,1]
e 5.5637832
x 0.9175763
```

Zur Kontrolle die automatische Lösung:

```
> coef(lm(y~x))
(Intercept)          x
5.5637832      0.9175763
```

Nun werden wir erklären, wie man zum System der Normalgleichungen kommt.

Die Normalgleichungen entstehen dadurch, dass die optimale Lösung des LSQ Problems zwei Bedingungen erfüllen muss:

1. Die Residuen müssen den Mittelwert Null haben, dh.

$$0 = \hat{\epsilon} \cdot \mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{e}$$

2. Die Residuen müssen zur Prädiktorvariablen unkorreliert sein, dh.

$$0 = \hat{\epsilon} \cdot (\mathbf{x} - \bar{x}\mathbf{e}) = (\mathbf{y} - \hat{\mathbf{y}}) \cdot (\mathbf{x} - \bar{x}\mathbf{e})$$

Wir erhalten so das Gleichungssystem

$$\begin{aligned} (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{e} = 0 &\implies \mathbf{y} \cdot \mathbf{e} = \hat{\mathbf{y}} \cdot \mathbf{e} \\ (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{x} = 0 &\implies \mathbf{y} \cdot \mathbf{x} = \hat{\mathbf{y}} \cdot \mathbf{x} \end{aligned}$$

Wir ersetzen nun

$$\hat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}$$

und erhalten

$$\begin{aligned} \mathbf{y} \cdot \mathbf{e} = \hat{\beta}_0 \mathbf{e} \cdot \mathbf{e} + \hat{\beta}_1 \mathbf{e} \cdot \mathbf{x} \\ \mathbf{y} \cdot \mathbf{x} = \hat{\beta}_0 \mathbf{e} \cdot \mathbf{x} + \hat{\beta}_1 \mathbf{x} \cdot \mathbf{x} \end{aligned} \implies \begin{pmatrix} \mathbf{y} \cdot \mathbf{e} \\ \mathbf{y} \cdot \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{e} \cdot \mathbf{e} & \mathbf{e} \cdot \mathbf{x} \\ \mathbf{e} \cdot \mathbf{x} & \mathbf{x} \cdot \mathbf{x} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

Das sind die Normalgleichungen, also jenes lineare Gleichungssystem, aus dem die Regressionsparameter berechnet werden können. Für uns ist von besonderem Interesse, wie die Koeffizienten dieses Gleichungssystems auf einfache Weise berechnet werden können.

Die linke Seite dieses Gleichungssystems läßt sich mit der Designmatrix  $\mathbf{X}$  ganz kompakt anschreiben als

$$\begin{pmatrix} \mathbf{y} \cdot \mathbf{e} \\ \mathbf{y} \cdot \mathbf{x} \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + \dots + y_n \\ x_1 y_1 + x_2 y_2 + \dots + x_n y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{X}^t \mathbf{y}$$

Aber auch die Matrix auf der rechten Seite des Gleichungssystems läßt sich durch die Designmatrix  $\mathbf{X}$  einfach darstellen:

$$\begin{pmatrix} \mathbf{e} \cdot \mathbf{e} & \mathbf{e} \cdot \mathbf{x} \\ \mathbf{e} \cdot \mathbf{x} & \mathbf{x} \cdot \mathbf{x} \end{pmatrix} = \mathbf{X}^t \mathbf{X}$$

Daher lautet das System der Normalgleichungen

$$\mathbf{X}^t \mathbf{y} = \mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}}$$

### 5.1.3 Beurteilung der Modellanpassung

Wenn man ein Regressionsmodell an die Daten angepasst hat (**model fitting**), dann stellt sich die Frage, ob die Prädiktorvariable  $X$  überhaupt einen Beitrag zur Erklärung der Responsevariablen  $Y$  liefert, und wie groß dieser Erklärungsbeitrag ist.

Es geht dabei um drei verschiedene Fragestellungen:

1. **Wie groß ist der Erklärungsbeitrag von  $X$  ?**
2. **Ist der Erklärungsbeitrag signifikant ?** Damit ist gemeint: Ist der Erklärungsbeitrag überzufällig, oder kann er in diesem Ausmaß auch durch Zufall allein zustande kommen ?
3. **Ist der Erklärungsbeitrag relevant ?** Damit ist gemeint: Ist der Erklärungsbeitrag groß genug, um im gegebenen Kontext inhaltlich von Bedeutung zu sein ?

Wir werden alle drei Fragen behandeln.

## Das Ausmaß des Erklärungsbeitrags

---

Das Ausmaß des Erklärungsbeitrags mißt man, indem man die Größe der Residuen gegenüberstellt, die entstehen, wenn man die Prädiktorvariable  $X$  einbezieht bzw. nicht einbezieht.

**Das Modell**  $M_0$ :  $y_i = \beta_0 + \epsilon_i$

Dieses Modell versucht, die Responsevariable ohne Berücksichtigung der Prädiktorvariablen  $X$  zu erklären. Wenn man dieses Modell mit LSQ anpasst, erhält man als optimalen Parameter den Mittelwert:

$$\hat{\beta}_0 = \bar{y}$$

Die optimalen Residuen lauten  $\hat{\epsilon}_i = y_i - \bar{y}$  und daher beträgt die Fehlerquadratsumme (Abweichungsquadratsumme) dieses Modells

$$AQS_{M_0} = \|\hat{\epsilon}\|^2 = \|\mathbf{y} - \bar{y}\mathbf{e}\|^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

**Das Modell**  $M_1$ :  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

Dieses Modell versucht, die Responsevariable unter Einbeziehung der Prädiktorvariablen  $X$  zu erklären. Wenn man dieses Modell anpasst, erhält man als optimale Parameter die Regressionsparameter  $\hat{\beta}_0$  und  $\hat{\beta}_1$ . Die Fehlerquadratsumme (Abweichungsquadratsumme) dieses Modells beträgt

$$AQS_{M_1} = \|\hat{\epsilon}\|^2 = \|\mathbf{y} - \hat{\beta}_0 - \hat{\beta}_1 \mathbf{x}\|^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

**Erklärungsbeitrag:**

Die Fehlerquadratsumme  $AQS_{M_1}$  des Modells  $M_1$  ist sicher kleiner als die Fehlerquadratsumme  $AQS_{M_0}$  des einfacheren Modells  $M_0$ :

$$AQS_{M_1} < AQS_{M_0}$$

Die Verminderung der Fehlerquadratsumme, die durch die Einbeziehung der Prädiktorvariablen  $X$  verursacht wird, nennen wir die **Devianz** des Modells  $M_1$  bzgl.  $M_0$  oder den **Erklärungsbeitrag** der Prädiktorvariablen  $X$  zur Erklärung von  $Y$ :

$$AQS_{M_0} - AQS_{M_1}$$

## Das Signifikanzproblem

---

Es gibt zwei Möglichkeiten, die Signifikanz des Erklärungsbeitrags der Prädiktorvariablen zu beurteilen. Der wichtigere und universellere Weg besteht darin, der Erklärungsbeitrag selbst in einer Varianzanalyse zu beurteilen. Eine andere Möglichkeit ist, die Streuung der Parameterschätzungen zu analysieren.

Wir beginnen mit der **Varianzanalyse** (ANOVA, Analysis of Variance).

Es geht darum eine statistischen Test durchzuführen, der eine Entscheidung zwischen den beiden Modellen trifft:

$$\begin{array}{ll} \text{Nullhypothese:} & M_0 \\ \text{Alternativhypothese:} & M_1 \end{array}$$

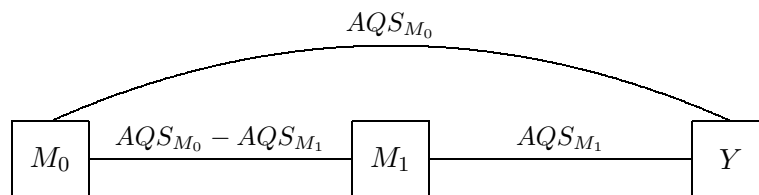
Der statistische Test soll entscheiden, ob das komplexere Modell  $M_1$  die Daten signifikant besser erklärt als das einfachere Modell  $M_0$ .

Die Methode der Varianzanalyse besteht darin, die Abweichungsquadratsummen der Modelle  $M_0$  und  $M_1$  gegenüberzustellen. Die Abweichungsquadratsumme des Modells  $M_0$  lässt sich in zwei Teile aufspalten:

$$\underbrace{AQS_{M_0}}_{\text{Gesamtstreuung } SS_T} = \underbrace{(AQS_{M_0} - AQS_{M_1})}_{\text{Erklärbare Streuung } SS^*} + \underbrace{AQS_{M_1}}_{\text{Reststreuung } SS_R}$$

Die in dieser Zerlegung (ANOVA) auftretenden Größen haben folgende Bedeutung:

$$\begin{aligned} SS_T &:= AQS_{M_0}: && \text{Die durch } M_0 \text{ nicht erklärbare Streuung von } Y \\ SS^* &:= AQS_{M_0} - AQS_{M_1}: && \text{Die durch } M_1 \text{ zusätzlich erklärbare Streuung von } Y \\ SS_R &:= AQS_{M_1}: && \text{Die durch } M_1 \text{ nicht erklärbare Streuung von } Y \end{aligned}$$



Wenn das Modell  $M_1$  keinen Erklärungswert besitzt (dh. wenn die Nullhypothese zutrifft), dann müssen die **mittleren Abweichungsquadratsummen**

$$MSS^* := \frac{SS^*}{1}, \quad MSS_R := \frac{SS_R}{n-2}$$

den gleichen Erwartungswert besitzen. Man bildet daher die sogenannte **F-Größe**

$$F := \frac{MSS^*}{MSS_R}$$

und beurteilt, ob sie signifikant vom Wert 1 abweicht. Die Beurteilung erfolgt durch ihren P-Wert (das ist die Wahrscheinlichkeit, mit der der aktuelle Wert der F-Größe auftreten kann, obwohl die Nullhypothese zutrifft).

Die für die Varianzanalyse relevanten Größen fasst man für gewöhnlich in einer Tabelle zusammen:

ANOVA-Tabelle

	Df	SumSq	MeanSq	F-Value	P-Value
X	1	$SS^* = AQS_{M_0} - AQS_{M_1}$	$MSS^*$	F	P
Residuals	$n - 2$	$SS_R = AQS_{M_1}$	$MSS_R$		

(5.7) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: CTWGT, Prädiktor: FBAG.

Führen Sie eine Varianzanalyse durch !

**Lösung:** Wir führen die Modellanpassung durch:

```
> fm=lm(CTWGT~FBAG)
```

Die Anzeige von **fm** liefert uns nur die Regressionskoeffizienten:

```
> fm
```

```
Call:
```

```
lm(formula = CTWGT ~ FBAG)
```

```
Coefficients:
```

```
(Intercept)          FBAG
      63.0811          0.2493
```

Für die ANOVA-Tabelle gibt es eine besondere Funktion:

```
> anova(fm)
```

```
Analysis of Variance Table
```

```
Response: CTWGT
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
FBAG         1   3569     3569  18.586 1.747e-05 ***
Residuals 1294 248517      192
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wir überzeugen uns durch direkte Berechnung der Abweichungsquadratsummen von der Richtigkeit der Größen in der ANOVA-Tabelle:

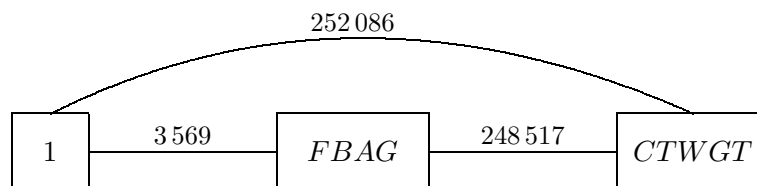
```
> norm(CTWGT-fitted(fm))^2
```

```
[1] 248517.1
```

```
> norm(fitted(fm)-mean(CTWGT))^2
```

```
[1] 3569.495
```

Wir können die Ergebnisse unserer Modellanpassung auch graphisch darstellen:




---

Wir beschreiben noch kurz die zweite Möglichkeit, das Signifikanzproblem der Modellanpassung zu behandeln. Obwohl diese zweite Methode keineswegs so informativ ist wie eine Varianzanalyse, wird sie doch sehr oft verwendet.

Der Grundgedanke besteht in folgender Überlegung.

Wenn das Modell  $M_1$ , dh. die Prädiktorvariable  $X$  keinen Erklärungsbeitrag für die Responsevariable  $Y$  liefert, dann müsste bei der Modellanpassung von  $M_1$  der Regressionsparameter  $\hat{\beta}_1$  gleich Null sein. Er ist natürlich nicht exakt gleich Null, sondern hat auf Grund der Zufälligkeit der Daten irgendeinen von Null verschiedenen Wert.

Wenn wir also bei der Modellanpassung von  $M_1$  auf einen von Null verschiedenen Wert von  $\hat{\beta}_1$  treffen, und das wird in der Regel immer der Fall sein, dann müssen wir entscheiden, ob dies nur durch die Zufälligkeit der Daten zustande gekommen ist, oder ob tatsächlich ein Erklärungsbeitrag der Prädiktorvariablen  $X$  dahinter steckt.

Man kann diese Signifikanzbeurteilung nun auch so durchführen, dass man den Regressionsparameter  $\hat{\beta}_1$  standardisiert, dh. durch seine Standardabweichung (SD) dividiert, und den entstehenden Quotienten (t-Value) auf Signifikanz überprüft (P-Value).

(5.8) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: CTWGT, Prädiktor: FBAG.

Führen Sie einen Signifikanztest für den Regressionsparameter durch !

**Lösung:** Die Funktion **summary** gibt eine Übersicht über alle wesentlichen Informationen in Zusammenhang mit der Modellanpassung:

```
> summary(fm)
```

Call:

```
lm(formula = CTWGT ~ FBAG)
```

Residuals:

Min	1Q	Median	3Q	Max
-29.048	-9.306	-2.308	6.444	72.688

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	63.08110	1.86204	33.877	< 2e-16 ***
FBAG	0.24925	0.05782	4.311	1.75e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.86 on 1294 degrees of freedom

Multiple R-Squared: 0.01416, Adjusted R-squared: 0.0134

F-statistic: 18.59 on 1 and 1294 DF, p-value: 1.747e-05

Sie enthält auch einige Zahlen, die wir noch nicht interpretieren können. Für uns ist jetzt nur der Signifikanztest für den Regressionsparameter  $\beta_1$  interessant. (Der gleichzeitig auch durchgeführte Signifikanztest für  $\beta_0$  hat für uns keine Bedeutung.)

Man kann die Frage stellen, wie man die Standardabweichung des Regressionskoeffizienten berechnet. Diese Frage kann aber nur mit Mitteln der Wahrscheinlichkeitstheorie beantwortet werden. Diese Mittel stehen uns hier nicht zu Verfügung.

## Das Relevanzproblem

Beim Relevanzproblem geht es um die Frage, ob der Erklärungsbeitrag der Prädiktorvariablen  $X$  (sofern er überhaupt signifikant ist) so groß ist, dass er inhaltlich von Bedeutung

ist. Zu dieser Frage können statistische Methoden natürlich nur die quantitativen Grundlagen liefern. Die Entscheidung, was inhaltlich als relevant anzusehen ist, kann nur aus dem sachlichen Kontext des Problems heraus entschieden werden.

Der Erklärungsbeitrag selbst, also die Größe

$$AQS_{M_0} - AQS_{M_1}$$

kann für sich genommen noch nicht als Maßzahl für die Relevanz verwendet werden, weil er skalenabhängig (maßstabsabhängig) ist. Eine skalenunabhängige Größe ist hingegen das **Bestimmtheitsmaß**

$$\frac{SS^*}{SS_R} = \frac{AQS_{M_0} - AQS_{M_1}}{AQS_{M_0}}$$

welches den relativen Erklärungsbeitrag (Prozentsatz), gemessen an der ursprünglichen Fehlerquadratsumme, angibt.

Eine andere Möglichkeit, die Relevanz skalenunabhängig zu messen, ist der Korrelationskoeffizient zwischen den Werten  $y_i$  der Responsevariablen und den Vorhersagewerten (fitted values)  $\hat{y}_i$ :

$$r_{y\hat{y}} = \frac{(\mathbf{y} - \bar{y}\mathbf{e}) \cdot (\hat{\mathbf{y}} - \bar{y}\mathbf{e})}{\|(\mathbf{y} - \bar{y}\mathbf{e})\| \|(\hat{\mathbf{y}} - \bar{y}\mathbf{e})\|}$$

Dieser Korrelationskoeffizient ist von großer Bedeutung für die Beurteilung von Regressionsproblemen. Allerdings wird diese Tatsache bei einem einfachen Regressionsproblem wie dem unseren noch nicht wirklich deutlich, denn hier gilt

$$r_{y\hat{y}} = r_{yx}$$

weil ja  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$  nur eine lineare Funktion von  $x_i$  ist.

Die beiden Maßzahlen für die Relevanz hängen auf einfache Weise miteinander zusammen:

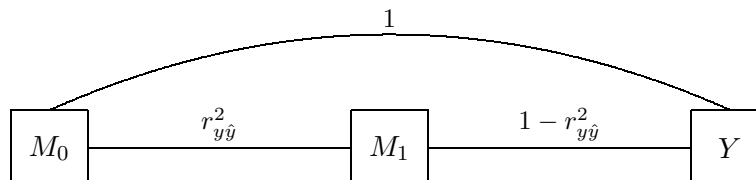
(5.9) SATZ:

*Das Bestimmtheitsmaß ist identisch mit RSquare, das heisst*

$$\frac{SS^*}{SS_R} = \frac{AQS_{M_0} - AQS_{M_1}}{AQS_{M_0}} = r_{y\hat{y}}^2$$

Auf Grund dieser für die Regressionstheorie grundlegenden Beziehung wird das Bestimmtheitsmaß meist als **RSquare** angegeben.

Wir können die Varianzanalyse statt mit den Abweichungssummen  $AQS$  mit den relativen Abweichungssummen durchführen. Das sieht dann so aus:



Mit unseren mathematischen Mitteln können wir die Gültigkeit der Gleichung leicht nachvollziehen.

Zur Berechnung der Größe der Fehlerquadratsumme verwenden wir unser Wissen über die Eigenschaften einer orthogonalen Projektion. Wenn wir einen Vektor  $\mathbf{a}$  auf die Richtung eines Vektors  $\mathbf{b}$  projizieren, dann gilt für den Rest  $\mathbf{r}$  der orthogonalen Zerlegung

$$\mathbf{a} = \lambda_{\text{opt}} \mathbf{b} + \mathbf{r}$$

die Gleichung

$$\|\mathbf{r}\|^2 = \|\mathbf{a}\|^2 (1 - c(\mathbf{a}, \mathbf{b})^2)$$

Diese Gleichung machen wir uns nun zunutze.

Die Modellgleichung  $y_i - \bar{y} = \hat{\beta}_1(x_i - \bar{x}) + \hat{\epsilon}_i$ , in Vektorschreibweise

$$\mathbf{y} - \bar{y}\mathbf{e} = \hat{\beta}_1(\mathbf{x} - \bar{x}\mathbf{e}) + \hat{\epsilon}$$

ist eine orthogonale Zerlegung und erfüllt daher

$$\|\hat{\epsilon}\|^2 = \|\mathbf{y} - \bar{y}\mathbf{e}\|^2 (1 - c(\mathbf{y} - \bar{y}\mathbf{e}, \mathbf{x} - \bar{x}\mathbf{e})^2)$$

In dem wir diese Gleichung in die statistische Bezeichnungsweise übersetzen, erhalten wir die Größe der Fehlerquadratsumme des Modells  $M_1$ :

$$AQS_{M_1} = AQS_{M_0} (1 - r_{yx}^2)$$

Diese Gleichung ist aber bereits die gewünschte Beziehung zwischen dem Bestimmtheitsmaß und RSquare.

(5.10) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: CTWGT, Prädiktor: FBAG.

Beurteilen Sie die Relevanz des Prädiktors für Responsevariable !

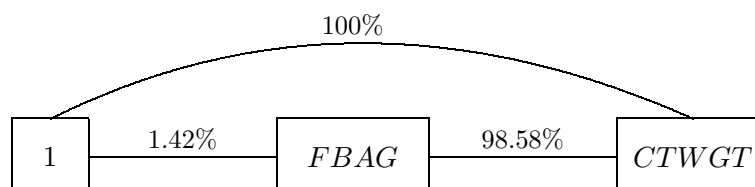
**Lösung:** Wir können das Bestimmtheitsmaß entweder als RSquare aus **summary(fm)** entnehmen, oder selbst berechnen aus

```
> aqs.0=norm(CTWGT-mean(CTWGT))^2
> aqs.1=norm(CTWGT-fitted(fm))^2
> (aqs.0-aqs.1)/aqs.0
[1] 0.0141598
```

oder

```
> r=cor(CTWGT,fitted(fm))
> r^2
[1] 0.0141598
```

Das Schema der Varianzanalyse mit relativen Größen sieht dann so aus:



## 5.1.4 Aufgaben

Die folgenden Aufgaben sollen mit quantitativen Variablen der Datensätze **statlab** und **kreditscoring** und einem Prädiktor durchgeführt werden.

(5.11) **Aufgabe:** Berechnen Sie die Regressionskoeffizienten mit Hilfe der R-Funktion **lm**.

(5.12) **Aufgabe:** Berechnen Sie die Regressionskoeffizienten mit Hilfe der expliziten Formeln.

(5.13) **Aufgabe:** Stellen Sie die Designmatrix auf und berechnen Sie die Regressionskoeffizienten mit Hilfe der Normalgleichungen.

(5.14) **Aufgabe:** Zeichnen Sie das Streudiagramm, zeichnen Sie die Regressionsgerade und die Residuen ein.

(5.15) **Aufgabe:** Berechnen Sie die Abweichungsquadratsummen für die ANOVA-Tabelle ohne die Funktionen **summary** und **anova** zu verwenden.

(5.16) **Aufgabe:** Stellen Sie ANOVA-Tabelle (Funktion **anova**) auf und interpretieren Sie sie hinsichtlich des Signifikanzproblems. Stellen Sie die Varianzanalyse schematisch als Diagramm dar.

(5.17) **Aufgabe:** Beurteilen Sie die Relevanz des Erklärungswerts.

(5.18) **Aufgabe:** Berechnen Sie das Bestimmtheitsmaß auf zwei Arten. Beschriften Sie das Schema der Varianzanalyse mit relativen Erklärungswerten.

### (5.19) TESTFRAGEN

1. Was ist ein einfaches lineares Regressionsmodell ? Erklären Sie die Begriffe Regressionsparameter und Residuen !
2. Was versteht man unter Modellanpassung mit dem LSQ-Prinzip ?
3. Worin besteht der Zusammenhang zwischen dem LSQ-Prinzip und orthogonalen Projektionen ? Begründen Sie damit die Formeln für die Regressionsparameter ?
4. Welche zwei Bedingungen an die Residuen eines optimal angepassten Modells erfüllen ?
5. Was ist die Designmatrix ? Was sind die Normalgleichungen und wozu dienen sie ?
6. Erklären Sie das Signifikanzproblem, das Relevanzproblem, und den Unterschied zwischen beiden !
7. Was ist die Devianz ? Wie ist sie definiert und was hat man sich darunter vorzustellen ?
8. Erklären Sie den Aufbau der ANOVA-Tabelle und die Bedeutung der darin enthaltenen Größen !
9. Wie beantwortet man das Signifikanzproblem ?
10. Wie beantwortet man das Relevanzproblem ?
11. Was ist RSquare ? Was ist das Bestimmtheitsmaß ? Wie hängen diese Begriffe zusammen ?

## 5.2 Das zweifache Regressionsmodell

### 5.2.1 Einführung

Beim zweifachen Regressionsmodell geht es darum, eine Responsevariable  $Y$  durch zwei Prädiktorvariable  $X_1$  und  $X_2$  zu erklären, dh. möglichst gut vorherzusagen.

Es liege eine Stichprobe vom Umfang  $n$  vor, also Daten  $y_1, y_2, \dots, y_n$  der Responsevariablen und Daten  $x_{11}, x_{21}, \dots, x_{n1}$  der Prädiktorvariablen  $X_1$  sowie Daten  $x_{12}, x_{22}, \dots, x_{n2}$  der Prädiktorvariablen  $X_2$ .

(5.20) DEFINITION:

*Unter einem zweifachen linearen Regressionsmodell versteht man ein Modell der Form*

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

*Dabei bezeichnen  $\beta_0$ ,  $\beta_1$  und  $\beta_2$  die Modellparameter und  $\epsilon_i$  die Residuen.*

Mit einem zweifachen Regressionsmodell versucht man, die Daten  $y_i$  durch die lineare Funktion  $\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$  vorherzusagen. Der Fehler, den man dabei macht, ist

$$y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} = \epsilon_i$$

Das sind die Residuen. Das Problem besteht daher wieder darin, ein optimales Modell zu finden, dh. ein Modell, bei dem die Residuen möglichst klein sind.

Die einfachste Methode, ein optimales Regressionsmodell zu finden, ist auch hier die **Methode der kleinsten Quadrate**, abgekürzt **LSQ**.

(5.21) DEFINITION:

*Unter dem **LSQ-Prinzip** versteht man die Optimierungsaufgabe*

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 = \text{Min} !$$

*Die Modellparameter sollen also so gewählt werden, dass die Quadratsumme der Residuen minimal ist.*

Die optimalen Modellparameter  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  und  $\hat{\beta}_2$  heißen **Regressionsparameter**. Die optimale Prognosefunktion

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$

ist die **lineare Regressionsfunktion**. Für die speziellen Residuen, die beim optimalen Modell anfallen, verwenden wir wieder ein besonderes Symbol:

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

## LSQ mit R

---

Auch ein zweifaches lineares Regressionsmodell wird in R durch eine Modellformel definiert: Man benennt die Responsevariable und trennt mit „~“ die darauffolgende Benennung der Prädiktorvariablen:  $y \sim x_1 + x_2$ . Die Prädiktorvariablen werden durch „+“ verbunden.

Zu einer solchen Modellformel mit mehr als einer einzigen Prädiktorvariablen sind zwei Bemerkungen zu machen.

**Erstens** hat die Verknüpfung der Prädiktorvariablen durch „+“ nichts mit der Rechenoperation der Addition zu tun. Wir werden später erfahren, was man tun muss, wenn man im Rahmen der Modellformel Berechnungen durchführen will.

**Zweitens** ist die Frage zu stellen, ob die **Reihenfolge**, mit der man die Prädiktorvariablen in der Modellformel angibt, eine Rolle spielt. Das ist ein Thema, das nicht ganz einfach mit ja oder nein zu beantworten ist. Wir werden im folgenden immer wieder darauf zurückkommen. Aber schon jetzt geben wir einen wichtigen Hinweis: Für die Schätzung der Regressionsparameter spielt die Reihenfolge keine Rolle, für die Beurteilung des Erklärungsbeitrags spielt sie jedoch eine ganz entscheidende Rolle.

Die Berechnung des optimalen Modells erfolgt durch die Funktion **lm** (linear model). Es ist wieder am besten, wenn man das Ergebnis der Funktion in einer Variablen (z.B. **fm**, fitted model) abspeichert, um beim Aufruf der Ausgabefunktionen das Modell nicht immer wieder neu berechnen zu müssen.

---

(5.22) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: CTWGT, Prädiktoren: MBAG, FBAG.

**Lösung:**

Wir berechnen das Modell:

```
> fm=lm(CTWGT~MBAG+FBAG)
```

Dadurch erhalten wir die Regressionskoeffizienten:

```
> coef(fm)
(Intercept)          MBAG          FBAG
61.97228442  0.23114630  0.07700842
```

Eine übersichtliche Grafik ist bei mehr als zwei beteiligten Variablen nicht mehr möglich. Wir können uns aber eine Übersicht über alle paarweisen Beziehungen verschaffen:

```
> pairs(cbind(CTWGT,MBAG,FBAG))
```

Wir sehen hier, dass zwar MBAG und FBAG stark korreliert sind, dass aber Korrelationen zwischen CTWGT und den Prädiktorvariablen grafisch allein nicht überzeugend nachgewiesen werden können.

---

## 5.2.2 Die mathematische Lösung

Wir sehen uns jetzt an, wie man zu den Lösungsformeln für die Regressionsparameter kommt. Wir beschränken uns dabei auf die Methode der Normalgleichungen.

### Die Normalgleichungen

---

(5.23) SATZ:

Das System der Normalgleichungen hat die Form

$$\mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{y}$$

wobei

$$\mathbf{X} := \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$$

die **Designmatrix** des linearen Regressionsmodells bezeichnet.

Bevor wir erklären, wie dieses System von Normalgleichungen entsteht, wollen wir es praktisch erproben.

---

(5.24) BEISPIEL:

Führen Sie mit dem Datensatz **statlab** eine Regressionsanalyse durch:

Response: CTWGT, Prädiktoren: MBAG, FBAG.

Berechnen Sie die Regressionsparameter mit dem System der Normalgleichungen.

**Lösung:** Wir berechnen zunächst die Designmatrix:

```
> x1=MBAG
> x2=FBAG
> y=CTWGT
> e=rep(1,1296)
> X=cbind(e,x1,x2)
```

Die Lösung der Normalgleichungen erhalten wir durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$$

also

```
> beta=solve(t(X)**X)**t(X)**y
> beta
      [,1]
e  61.97228442
x1  0.23114630
x2  0.07700842
```

Zur Kontrolle nochmals die automatische Lösung:

```
> coef(fm)
(Intercept)          MBAG          FBAG
61.97228442  0.23114630  0.07700842
```

---

Nun werden wir erklären, wie man zum System der Normalgleichungen kommt.

Die Normalgleichungen entstehen immer dadurch, dass die optimale Lösung des LSQ Problems zwei Bedingungen erfüllen muss:

1. **Die Residuen müssen den Mittelwert Null haben**, dh.

$$0 = \hat{\mathbf{e}} \cdot \mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{e}$$

2. **Die Residuen müssen zu allen Prädiktorvariablen unkorreliert sein**, dh.

$$0 = \hat{\mathbf{e}} \cdot (\mathbf{x}_1 - \bar{x}_1 \mathbf{e}) = (\mathbf{y} - \hat{\mathbf{y}}) \cdot (\mathbf{x}_1 - \bar{x}_1 \mathbf{e})$$

$$0 = \hat{\mathbf{e}} \cdot (\mathbf{x}_2 - \bar{x}_2 \mathbf{e}) = (\mathbf{y} - \hat{\mathbf{y}}) \cdot (\mathbf{x}_2 - \bar{x}_2 \mathbf{e})$$

Wir erhalten so das Gleichungssystem

$$(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{e} = 0 \quad \implies \quad \mathbf{y} \cdot \mathbf{e} = \hat{\mathbf{y}} \cdot \mathbf{e}$$

$$(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{x}_1 = 0 \quad \implies \quad \mathbf{y} \cdot \mathbf{x}_1 = \hat{\mathbf{y}} \cdot \mathbf{x}_1$$

$$(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{x}_2 = 0 \quad \implies \quad \mathbf{y} \cdot \mathbf{x}_2 = \hat{\mathbf{y}} \cdot \mathbf{x}_2$$

Wir ersetzen nun

$$\hat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_1 + \hat{\beta}_2 \mathbf{x}_2$$

und erhalten

$$\begin{aligned} \mathbf{y} \cdot \mathbf{e} &= \hat{\beta}_0 \mathbf{e} \cdot \mathbf{e} + \hat{\beta}_1 \mathbf{x}_1 \cdot \mathbf{e} + \hat{\beta}_2 \mathbf{x}_2 \cdot \mathbf{e} \\ \mathbf{y} \cdot \mathbf{x}_1 &= \hat{\beta}_0 \mathbf{e} \cdot \mathbf{x}_1 + \hat{\beta}_1 \mathbf{x}_1 \cdot \mathbf{x}_1 + \hat{\beta}_2 \mathbf{x}_2 \cdot \mathbf{x}_1 \\ \mathbf{y} \cdot \mathbf{x}_2 &= \hat{\beta}_0 \mathbf{e} \cdot \mathbf{x}_2 + \hat{\beta}_1 \mathbf{x}_1 \cdot \mathbf{x}_2 + \hat{\beta}_2 \mathbf{x}_2 \cdot \mathbf{x}_2 \end{aligned} \quad \implies \quad \begin{pmatrix} \mathbf{y} \cdot \mathbf{e} \\ \mathbf{y} \cdot \mathbf{x}_1 \\ \mathbf{y} \cdot \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{e} \cdot \mathbf{e} & \mathbf{x}_1 \cdot \mathbf{e} & \mathbf{x}_2 \cdot \mathbf{e} \\ \mathbf{e} \cdot \mathbf{x}_1 & \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_1 \\ \mathbf{e} \cdot \mathbf{x}_2 & \mathbf{x}_1 \cdot \mathbf{x}_2 & \mathbf{x}_2 \cdot \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

Das sind nun die Normalgleichungen, also jenes lineare Gleichungssystem, aus dem die Regressionsparameter berechnet werden können. Für uns ist wieder von besonderem Interesse, wie die Koeffizienten dieses Gleichungssystems auf einfache Weise berechnet werden können.

Die linke Seite dieses Gleichungssystems lässt sich mit der Designmatrix  $\mathbf{X}$  ganz kompakt anschreiben als

$$\begin{pmatrix} \mathbf{y} \cdot \mathbf{e} \\ \mathbf{y} \cdot \mathbf{x}_1 \\ \mathbf{y} \cdot \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + \dots + y_n \\ x_{11}y_1 + x_{21}y_2 + \dots + x_{n1}y_n \\ x_{12}y_1 + x_{22}y_2 + \dots + x_{n2}y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{X}^t \mathbf{y}$$

Aber auch die Matrix auf der rechten Seite des Gleichungssystems lässt sich durch die Designmatrix  $\mathbf{X}$  einfach darstellen:

$$\begin{pmatrix} \mathbf{e} \cdot \mathbf{e} & \mathbf{x}_1 \cdot \mathbf{e} & \mathbf{x}_2 \cdot \mathbf{e} \\ \mathbf{e} \cdot \mathbf{x}_1 & \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_1 \\ \mathbf{e} \cdot \mathbf{x}_2 & \mathbf{x}_1 \cdot \mathbf{x}_2 & \mathbf{x}_2 \cdot \mathbf{x}_2 \end{pmatrix} = \mathbf{X}^t \mathbf{X}$$

Daher lautet das System der Normalgleichungen

$$\mathbf{X}^t \mathbf{y} = \mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}}$$

### 5.2.3 Die Beurteilung der Modellanpassung

Bei einem linearen Regressionsmodell mit mehr als einer Prädiktorvariablen ist die Beurteilung der Modellanpassung wesentlich komplizierter als bei einem einfachen Regressionsmodell. Aber alle wichtigen Komplikationen treten schon beim zweifachen Regressionsmodell auf.

Um die Übersicht zu behalten, führen wir einige Bezeichnungen ein.

Variable	Symbol	Modellformel	Abweichungsquadratsumme
1	$M_0$	$Y \sim 1$	$AQS_0$
1, $X_1$	$M_1$	$Y \sim 1 + X_1$	$AQS_1$
1, $X_2$	$M_2$	$Y \sim 1 + X_2$	$AQS_2$
1, $X_1, X_2$	$M_{12}$	$Y \sim 1 + X_1 + X_2$	$AQS_{12}$

Die erwähnte Komplexität rührt daher, dass es nun nicht nur zwei Erklärungsmodelle gibt, wie bei der einfachen Regression, sondern vier davon. Wenn man nun von dem Erklärungsbeitrag des zweifachen Regressionsmodells  $M_{12}$  spricht, dann muss man dazu sagen, in bezug auf welches Ausgangsmodell dieser Erklärungsbeitrag gemeint ist.

## RSquare

Die Erklärungsbeiträge werden wir wie bisher als Differenzen der Abweichungsquadratsummen definieren. Allerdings wollen wir gleich von Beginn an jene Vereinfachung vornehmen, die wir beim einfachen Regressionsmodell erst in Zusammenhang mit dem Relevanzproblem kennengelernt haben.

Die Abweichungsquadratsummen sind ja skalenabhängige Größen. Wir haben aber gesehen, dass man Erklärungsbeiträge auch durch Korrelationskoeffizienten beschreiben kann, und dass es einen ganz einfachen Zusammenhang zwischen diesen Größen gibt.

Es sei  $M$  irgendein lineares Erklärungsmodell. Wir bezeichnen mit  $\hat{Y}$  die Vorhersagewerte für  $Y$ , die dieses Modell nach einer LSQ-Anpassung liefert.

(5.25) DEFINITION:

*Unter dem multiplen Korrelationskoeffizienten des Modell  $M$  versteht man den Korrelationskoeffizienten*

$$R_{y.M} := \text{Corr}(Y, \hat{Y}),$$

*also den Korrelationskoeffizienten zwischen den Daten von  $Y$  und den Vorhersagewerten  $\hat{Y}$ .*

Das Vorzeichen des multiplen Korrelationskoeffizienten ist immer positiv, weil durch die Parameterschätzung immer eine positive Kopplung zwischen den Vorhersagewerten und den Daten erzwungen wird. Daher ist nur der Betrag des multiplen Korrelationskoeffizienten von Interesse.

Hier gilt nun die fundamentale Gleichung:

$$AQ S_0 - AQ S_M = AQ S_0 R_{y.M}^2$$

Also ist der Erklärungsbeitrag eines linearen Modells immer proportional zum Quadrat des multiplen Korrelationskoeffizienten. Der gemeinsame Faktor  $AQ S_0$  enthält keine Information, und hängt nur von der verwendeten Meßskala ab. Wenn wir durch  $AQ S_0$  dividieren, dann erhalten wir die Gleichung

$$\frac{AQ S_0 - AQ S_M}{AQ S_0} = R_{y.M}^2 = \text{Bestimmtheitsmaß}$$

Das Bestimmtheitsmaß ist eine skalenunabhängige Maßzahl für den gesamten Erklärungswert eines Modell. Es gibt den Prozentsatz der durch das Modell  $M$  erklärten Streuung von  $Y$  an, gemessen an der Gesamtstreuung von  $Y$ .

Wir ergänzen unsere Tabelle mit Bezeichnungen:

Variable	Symbol	Modellformel	AQS	RSquare
1	$M_0$	$Y \sim 1$	$AQ S_0$	
1, $X_1$	$M_1$	$Y \sim 1 + X_1$	$AQ S_1$	$R_{y.x_1}^2$
1, $X_2$	$M_2$	$Y \sim 1 + X_2$	$AQ S_2$	$R_{y.x_2}^2$
1, $X_1, X_2$	$M_{12}$	$Y \sim 1 + X_1 + X_2$	$AQ S_{12}$	$R_{y.x_1x_2}^2$

## ANOVA-Zerlegungen

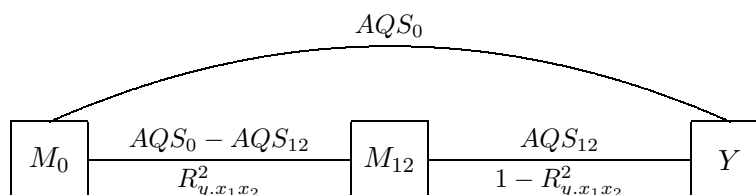
ANOVA:  $M_0 \subseteq M_{12}$

Wir beginnen damit, die Modelle  $M_0$  und  $M_{12}$  miteinander zu vergleichen. Die Varianzanalyse ergibt:

$$\begin{aligned} AQ S_0 &= (AQ S_0 - AQ S_{12}) + AQ S_{12} \\ 1 &= R_{y.x_1x_2}^2 + (1 - R_{y.x_1x_2}^2) \end{aligned}$$

Dabei bedeuten

$AQ S_0$  bzw. : Der Beitrag von 1 zur Erklärung von  $Y$   
 $AQ S_0 - AQ S_{12}$  bzw.  $R_{y.x_1x_2}^2$  : Der gemeinsame Beitrag von  $X_1$  und  $X_2$  über 1 hinaus  
 $AQ S_{12}$  bzw.  $1 - R_{y.x_1x_2}^2$  : Der unerklärte Rest



(5.26) BEISPIEL:

Datensatz **statlab**,  $Y = CBWGT$ ,  $X_1 = MBAG$ ,  $X_2 = FBAG$ .

Wir passen die Modelle  $M_0$  und  $M_{12}$  an und führen eine ANOVA durch:

```
> fm.0=lm(CTWGT~1)
> fm.12=lm(CTWGT~MBAG+FBAG)
> anova(fm.0, fm.12)
Analysis of Variance Table
```

Model 1: CTWGT ~ 1

Model 2: CTWGT ~ MBAG + FBAG

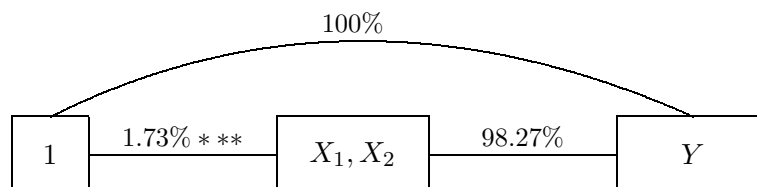
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	252087				
2	1293	247729	2	4358	11.373	1.27e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Diese Tabelle liefert uns die Abweichungsquadratsummen der beiden Modelle und ihre Differenz, den gewünschten Erklärungswert. Wir lesen daraus die Zunahme ab, die durch das Modell  $M_{12}$  erklärt werden:

	Zunahme von	
	Erklärungsbeitrag	RSquare
$M_{12}$	4 358	0.0173
Rest	247 729	0.9827
Summe	252 087	1



Der P-Wert der F-Größe sagt uns, dass der Erklärungsbeitrag signifikant ist. Das heisst in diesem Fall genau folgendes: Die Nullhypothese, dass  $X_1$  und  $X_2$  gemeinsam keinen Erklärungsbeitrag leisten, wird verworfen.

Man kann aber auch eine Folge von Modellen untersuchen, die immer komplexer werden.

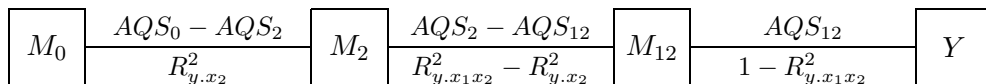
**ANOVA:**  $M_0 \subseteq M_2 \subseteq M_{12}$

Wir nehmen zunächst die Variable  $X_2$  ins Modell auf und anschließend  $X_1$ .

$$\begin{aligned} AQS_0 &= (AQS_0 - AQS_2) + (AQS_2 - AQS_{12}) + AQS_{12} \\ 1 &= R_{y.x_2}^2 + (R_{y.x_1x_2}^2 - R_{y.x_2}^2) + (1 - R_{y.x_1x_2}^2) \end{aligned}$$

Dabei bedeuten

$AQS_0$  bzw. : Der Beitrag von 1 zur Erklärung von  $Y$   
 $AQS_0 - AQS_2$  bzw.  $R_{y.x_2}^2$  : Der Beitrag von  $X_2$  über 1 hinaus  
 $AQS_2 - AQS_{12}$  bzw.  $R_{y.x_1x_2}^2 - R_{y.x_2}^2$  : Der Beitrag von  $X_1$  über 1 und  $X_2$  hinaus  
 $AQS_{12}$  bzw.  $1 - R_{y.x_1x_2}^2$  : Der unerklärte Rest



(5.27) BEISPIEL:

Datensatz **statlab**,  $Y = CBWGT$ ,  $X_1 = MBAG$ ,  $X_2 = FBAG$ .

Wir passen die Modelle  $M_0$ ,  $M_2$  und  $M_{12}$  an und führen eine ANOVA durch:

```

> fm.0=lm(CTWGT~1)
> fm.2=lm(CTWGT~FBAG)
> fm.12=lm(CTWGT~MBAG+FBAG)
> anova(fm.0, fm.2, fm.12)
  
```

Analysis of Variance Table

Model 1: CTWGT ~ 1

Model 2: CTWGT ~ FBAG

Model 3: CTWGT ~ MBAG + FBAG

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	252087				
2	1294	248517	1	3569	18.6307	1.707e-05 ***
3	1293	247729	1	788	4.1151	0.04271 *

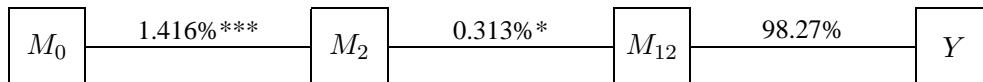
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Diese Tabelle liefert uns die Abweichungsquadratsummen der beiden Modelle und ihre

Differenz, den gewünschten Erklärungswert. Wir lesen daraus ab:

	Zunahme von	
	Erklärungswert	RSquare
$M_2$	3 569	0.01416
$M_{12}$	788	0.00313
Rest	247 729	0.98271
Summe	252 087	1



Die P-Werte der F-Größen liefern uns wieder die Informationen über die Signifikanz.

Der Erklärungsbeitrag des Prädiktors FBAG ist hoch signifikant, wenn er als erstes und alleine ins Modell eingebracht wird. Der zusätzliche Erklärungsbeitrag des Prädiktors MBAG ist wesentlich geringer, aber gerade noch signifikant.

Was passiert, wenn wir die Variablen  $X_1$  und  $X_2$  in umgekehrter Reihenfolge ins Modell einbringen ?

(5.28) BEISPIEL:

Datensatz **statlab**,  $Y = CBWGT$ ,  $X_1 = MBAG$ ,  $X_2 = FBAG$ .

Wir passen die Modelle  $M_0$ ,  $M_1$  und  $M_{12}$  an und führen eine ANOVA durch:

```

> fm.0=lm(CTWGT~1)
> fm.1=lm(CTWGT~MBAG)
> fm.12=lm(CTWGT~MBAG+FBAG)
> anova(fm.0, fm.1, fm.12)
Analysis of Variance Table
  
```

Model 1: CTWGT ~ 1

Model 2: CTWGT ~ MBAG

Model 3: CTWGT ~ MBAG + FBAG

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	252087				
2	1294	247836	1	4250	22.1833	2.745e-06 ***
3	1293	247729	1	108	0.5624	0.4534

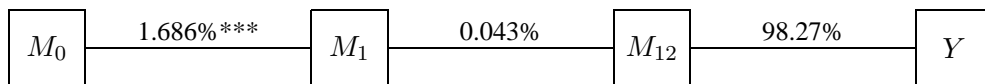
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Diese Tabelle liefert uns wieder die Abweichungsquadratsummen der beiden Modelle und

ihre Differenz, den gewünschten Erklärungswert. Wir lesen daraus ab:

	Zunahme von	
	Erklärungswert	RSquare
$M_1$	4 250	0.01686
$M_{12}$	108	0.00043
Rest	247 729	0.98271
Summe	252 087	1



Wenn wir hier die Signifikanz der Erklärungsbeiträge untersuchen ergibt sich ein ganz anderes Bild. Jetzt ist es der Prädiktor MBAG, der hochsignifikant ist, wenn er als erstes und alleine ins Modell eingebracht wird. Dagegen liefert FBAG keinen signifikanten zusätzlichen Erklärungsbeitrag.

## Modellselektion

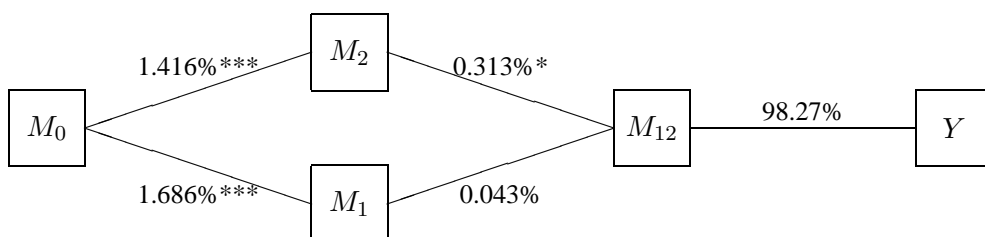
Das Modell  $M_{12}$  nennt man das **saturierte Modell**, weil in diesem Modell alle Prädiktorvariablen berücksichtigt sind, die überhaupt in betracht gezogen werden. Wenn der Erklärungsbeitrag des saturierten Modells signifikant ist, dann bedeutet das, dass es zumindest einige Prädiktorvariablen in diesem Modell gibt, die einen signifikanten Erklärungsbeitrag besitzen.

Es ist aber durchaus möglich und häufig der Fall, dass keineswegs alle Prädiktoren, die zu Beginn in betracht gezogen wurden, auch tatsächlich wirklich nötig sind. Natürlich hat von allen denkbaren Modellen das saturierte Modell den höchsten Erklärungsbeitrag. Aber ein einfacheres Modell, dh. ein Modell mit weniger Prädiktoren ist dann vorzuziehen, wenn der zusätzliche Erklärungsbeitrag des saturierten Modells nicht signifikant ist.

(5.29) BEISPIEL:

Datensatz **statlab**,  $Y = CBWGT$ ,  $X_1 = MBAG$ ,  $X_2 = FBAG$ .

Wir passen die Modelle  $M_0$ ,  $M_1$  und  $M_{12}$  an und wollen eine Modellselektion durchführen. Zu diesem Zweck sehen wir uns zunächst die Verhältnisse in einem Diagramm an:



Dieses Diagramm zeigt uns die Zunahmen an Erklärungswert, die durch die Aufnahme der Prädiktorvariablen in das Modell verursacht werden. Wir sehen noch einmal, dass beide

Erklärungsvariablen hoch signifikant sind, wenn sie alleine zur Erklärung herangezogen werden. Andererseits ist der zusätzliche Erklärungswert der Prädiktorvariablen eher gering. Nur MBAG liefert einen schwach signifikanten zusätzlichen Erklärungsbeitrag.

Wir wenden uns nun der Modellselektion zu.

Das saturierte Modell ist, wie wir wissen, signifikant. Wenn wir das saturierte Modell nun vereinfachen wollen, müssen wir solche Prädiktoren aus dem saturierten Modell entfernen, die keinen signifikanten Beitrag leisten.

Im vorliegenden Beispiel ist es so, dass der Prädiktor FBAG im saturierten Modell keinen signifikanten Beitrag leistet. Daher entfernen wir ihn aus dem Modell. Auf diese Weise gelangen wir zum Modell  $M_1$  mit dem Prädiktor MBAG. Dieses Modell läßt sich nicht weiter vereinfachen, denn der Erklärungsbeitrag von MBAG im Modell  $M_1$  ist signifikant.

Wir haben also den Prädiktor FBAG aus dem Modell entfernt, obwohl er signifikant ist, wenn er als erster und einziger Prädiktor ins Modell aufgenommen wird. Hätten wir den Prädiktor MBAG nicht zu Verfügung, dann wäre das Modell  $M_2$  durchaus ein brauchbares Erklärungsmodell. Da aber MBAG offenbar mehr erklärt, und weil FBAG nichts zusätzlich beiträgt, führt die Modellselektion zum optimalen Erklärungsmodell  $M_1$ .

(5.30) BEISPIEL:

Datensatz **statlab**,  $Y = CBWGT$ ,  $X_1 = MBAG$ ,  $X_2 = FBAG$ .

Wir passen die Modelle  $M_0$ ,  $M_1$  und  $M_{12}$  an und wollen eine Modellselektion durchführen.

Zunächst stellen wir fest, dass das saturierte Modell signifikant ist:

```
> anova(fm.0, fm.12)
Analysis of Variance Table

Model 1: FIT ~ 1
Model 2: FIT ~ MBAG + FBAG
  Res.Df    RSS   Df Sum of Sq    F    Pr(>F)
1     1295 6031071
2     1293 5911695    2    119376 13.055 2.437e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anschließend verschaffen wir uns die Details für die Modellselektion.

```
> fm.12=lm(FIT~MBAG+FBAG)
> fm.1=lm(FIT~MBAG)
> fm.2=lm(FIT~FBAG)
> fm.0=lm(FIT~1)

> anova(fm.0, fm.1, fm.12)
Analysis of Variance Table

Model 1: FIT ~ 1
Model 2: FIT ~ MBAG
Model 3: FIT ~ MBAG + FBAG
  Res.Df    RSS   Df Sum of Sq    F    Pr(>F)
```

```

1  1295 6031071
2  1294 6015831    1    15240  3.3332  0.06813 .
3  1293 5911695    1    104137 22.7767 2.027e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

> anova(fm.0, fm.2, fm.12)
Analysis of Variance Table

```

```
Model 1: FIT ~ 1
```

```
Model 2: FIT ~ FBAG
```

```
Model 3: FIT ~ MBAG + FBAG
```

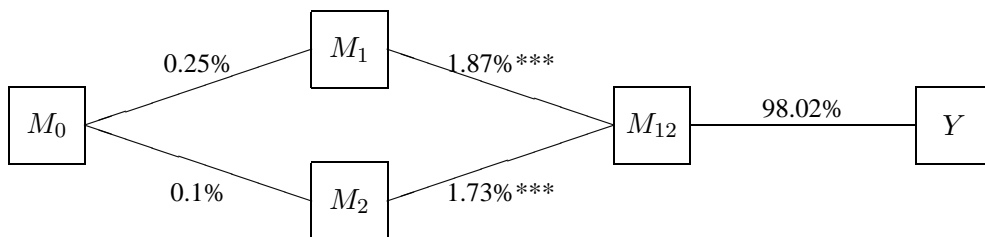
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	6031071				
2	1294	6024767	1	6304	1.3788	0.2405
3	1293	5911695	1	113072	24.7311	7.477e-07 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Wir bauen das Diagramm auf:



Dieses Ergebnis ist sehr interessant, weil es sich vom vorangehenden Beispiel völlig unterscheidet.

Keiner der beiden Prädiktoren leistet für sich allein einen signifikanten Erklärungsbeitrag. Gemeinsam ist jedoch der Erklärungsbeitrag hoch signifikant. Im saturierten Modell kann keiner der Prädiktoren weggelassen werden.

Um eine Interpretation dieses Ergebnisses zu erhalten, sehen wir uns die Regressionskoeffizienten an:

```

> coef(fm.12)
(Intercept)      MBAG      FBAG
 152.353296    2.768139   -2.393962

```

Sie sind ungefähr gleich groß, aber haben unterschiedliches Vorzeichen. Offenbar ist es die Differenz der Prädiktoren (Altersunterschied der Eltern), die die Responsevariable FIT erklärt: Je größer FBAG-MBAG ist, desto kleiner ist FIT !

An dieser Stelle muss allerdings ein Warnhinweis gegeben werden. Es ist keinesfalls so, dass ein solches Erklärungsmodell als kausales Modell interpretiert werden darf. Das geringere Familieneinkommen ist keinesfalls die Folge des Altersunterschieds der Eltern ! Vielmehr muss man hier vermuten, dass soziale und ethnische Faktoren sowohl die Variable FIT als auch die Differenz FBAG-MBAG beeinflussen.

Einen Hinweis gibt uns das Streudiagramm:

```

> plot(FBAG-MBAG, FIT)

```

Erst ab einem Altersunterschied von >10 Jahren scheinen die hohen Familieneinkommen zu fehlen. Man müßte genau untersuchen, welche Faktoren für diese Gruppe kennzeichnend sind. Es könnte sich z.B. um geschiedene und wiederverheiratete Männer handeln, deren FIT durch Alimentationszahlungen geringer ist.

---

Wir fassen zusammen:

- Als erstes ist zu prüfen, ob das saturierte Modell signifikant ist.  
Wenn ja, dann startet die Modellselektion.  
Wenn nein, dann liefern die Prädiktoren (einzeln und in ihrer Gesamtheit) keinen Erklärungsbeitrag.
- Die Modellselektion sucht ein einfacheres Modell als das saturierte Modell.
- Der Übergang zu einem einfacheren Modell ist dann möglich, wenn das saturierte Modell keinen darüber hinausgehenden signifikanten Erklärungsbeitrag leistet.

## Signifikanzprüfung der Parameter

---

Ein häufiger Fehler bei der Interpretation von Regressionsmodellen besteht darin, dass die Modellselektion nur auf Grund der Parameterschätzungen durchgeführt wird.

Die P-Werte der F-Größen in der Parameterschätzung betreffen aber lediglich die zusätzlichen Erklärungswerte der Prädiktoren, wenn sie jeweils als letzte ins Modell eingebracht werden. **Es ist daher falsch, all jene Prädiktoren aus dem Modell zu entfernen, deren Regressionskoeffizienten nicht signifikante P-Werte besitzen !**

---

(5.31) BEISPIEL:

Wir konstruieren ein Beispiel, an dem sehr drastisch deutlich wird, wie man es nicht machen darf.

```
> Z=rnorm(100)
> Y=Z+rnorm(100)
> X1=Z+0.01*rnorm(100)
> X2=Z+0.01*rnorm(100)
```

Dieses Datenmaterial ist so konstruiert, dass beide Prädiktoren  $X_1$  und  $X_2$  die Responsevariable  $Y$  sehr gut erklären, aber dass sie selbst sehr ähnlich sind, dh. hoch korreliert sind. Wenn man daher einen der beiden Prädiktoren ins Modell einführt, dann braucht man den jeweils anderen nicht mehr.

Die Parameterschätzung für das saturierte Modelleergibt:

```
> fm.12=lm(Y~X1+X2)
> summary(fm.12)
```

Call:

```
lm(formula = Y ~ X1 + X2)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
```

```
-2.90018 -0.57157 0.02383 0.61880 1.89941
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.19005    0.09329  -2.037  0.0444 *
X1           8.59120    6.25289   1.374  0.1726
X2          -7.60747    6.26229  -1.215  0.2274
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.9269 on 97 degrees of freedom
Multiple R-Squared: 0.572,      Adjusted R-squared: 0.5632
F-statistic: 64.82 on 2 and 97 DF,  p-value: < 2.2e-16
```

Das saturierte Modell ist also hoch signifikant. Aber beide Prädiktoren haben nicht signifikante Regressionsparameter.

Das ist aber nur scheinbar ein Widerspruch. Jeder der beiden Regressionsparameter kann für sich gleich Null gesetzt werden, weil ja der andere noch da ist. Aber man kann keinesfalls beide Prädiktoren aus dem Modell streichen !

Die ANOVA-Tabellen zeigen dies sehr deutlich:

```
> fm.1=lm(Y~X1)
> fm.2=lm(Y~X2)
> fm.0=lm(Y~1)
> anova(fm.0, fm.1, fm.12)
Analysis of Variance Table
```

```
Model 1: Y ~ 1
Model 2: Y ~ X1
Model 3: Y ~ X1 + X2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      99 194.725
2      98  84.611  1  110.114 128.1574 <2e-16 ***
3      97  83.343  1    1.268  1.4758 0.2274
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(fm.0, fm.2, fm.12)
Analysis of Variance Table
```

```
Model 1: Y ~ 1
Model 2: Y ~ X2
Model 3: Y ~ X1 + X2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      99 194.725
2      98  84.965  1  109.760 127.7454 <2e-16 ***
3      97  83.343  1    1.622  1.8878 0.1726
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Durchführung der Modellselektion

---

Es ist ziemlich aufwendig, alle möglichen ANOVA-Sequenzen aufzustellen, um alle nötigen AQS-Werte zu erhalten. Daher liegt es nahe, mit einem einfachen Programm die wichtigen Größen mit einem Schlag zu berechnen.

Zu diesem Zweck enthält unsere Library **LinAlg.r** die Funktion **modelselection**.

---

(5.32) BEISPIEL:

Wir beginnen mit dem Beispiel, das wir mittlerweile am besten kennen. Die Funktion **modelselection** wird auf eine Matrix angewendet, die als erste Spalte die Responsevariable, und in den anderen Spalten die Prädiktorvariablen enthält.

Nach Aufruf der Funktion werden für alle Modelle (alle denkbaren Teilmengen von Prädiktorvariablen) das Bestimmtheitsmaß und der zugehörige P-Wert berechnet.

Das sieht dann so aus:

```
> modelselection(cbind(CTWGT, MBAG, FBAG))
  ExplVar RSquare  PValue MBAG FBAG
1      2 0.01729          1     1
2      1 0.01416 0.04271     0     1
3      1 0.01686 0.45342     1     0
4      0          1e-05     0     0
```

---

Hierzu sind allerdings einige Erklärungen nötig.

Die Modelle werden in den letzten beiden Spalten beschrieben, und zwar durch Indikatoren (0 oder 1) der ins Modell genommenen Prädiktoren. Die Tabelle beginnt also mit dem saturierten Modell und endet mit dem Konstantenmodell. In der ersten Spalte wird angezeigt, wieviele Prädiktoren das jeweilige Modell enthält.

In der zweiten Spalte ist das Bestimmtheitsmaß (RSquare) des jeweiligen Modells angegeben. Es handelt sich also um den Anteil der durch das Modell erklärten AQS relativ zur AQS des Konstantenmodells.

Die dritte Spalte enthält die für die Modellselektion wesentliche Information. Es handelt sich um die P-Werte jener Nullhypothesen, die durch die letzten beiden Spalten definiert werden.

**Nullhypothese:** Die durch 0 gekennzeichneten Prädiktoren haben keinen Erklärungswert.

**P-Wert:** Wahrscheinlichkeit der beobachteten Daten unter der Voraussetzung, dass die Nullhypothese zutrifft.

Wenn wir nun für unser Beispiel die Modellselektion nach dieser Tabelle durchführen wollen, gehen wir so vor:

- Wir überprüfen zunächst, ob das saturierte Modell überhaupt signifikant ist. Zu diesem Zweck verwenden wir den P-Wert des Konstantenmodells. Ist dieser P-Wert kleiner als das von uns zugrundegelegte Signifikanzniveau (z.B.  $< 0.05$ ), dann wird das Konstantenmodell als Nullhypothese abgelehnt und daher liefert das saturierte Modell einen signifikanten Erklärungsbeitrag.

- Als nächsten Schritt werden alle weiteren Modelle ausgeschieden, die als Nullhypothese abgelehnt werden, dh. deren P-Wert kleiner ist als das von uns zugrundegelegte Signifikanzniveau (z.B.  $< 0.05$ ).

Alle übrigen Modelle sind mit dem Datenmaterial verträglich, dh. sie verlieren relativ zum saturierten Modell keine signifikante Information.

- Von allen Modellen, die verblieben sind, suchen wir das einfachste, dh. jenes mit der geringsten Anzahl von Prädiktoren.
- Falls es mehrere Modelle gibt, deren P-Wert nicht signifikant ist, und deren Anzahl von Prädiktoren minimal ist, dann wählen wir jenes Modell, das am besten zu den Daten passt, dh. jenes mit dem größten P-Wert.

#### (5.33) BEISPIEL:

Unser Beispiel

```
> modelselection(cbind(CTWGT, MBAG, FBAG))
  ExplVar RSquare  PValue MBAG FBAG
1         2 0.01729          1    1
2         1 0.01416 0.04271    0    1
3         1 0.01686 0.45342    1    0
4         0          1e-05    0    0
```

wird dann folgendermaßen interpretiert:

Das saturierte Modell ist signifikant, da Modell 4 abgelehnt wird. Modell 2 wird ebenfalls abgelehnt. Dagegen ist Modell 3 mit den Daten verträglich: Wir entscheiden uns für MBAG als alleinigen Prädiktor.

#### (5.34) BEISPIEL:

Das Beispiel

```
> modelselection(cbind(FIT, MBAG, FBAG))
  ExplVar RSquare  PValue MBAG FBAG
1         2 0.01979          1    1
2         1 0.00105          0    1
3         1 0.00253          0    0
4         0          0          0    0
```

wir so interpretiert:

Das saturierte Modell ist signifikant, aber keines der einfacheren Modelle ist mit den Daten verträglich. Daher entscheiden wir uns für das saturierte Modell.

#### (5.35) BEISPIEL:

Das Beispiel

```
> modelselection(cbind(Y,X1,X2))
  ExplVar RSquare  PValue X1 X2
1         2   0.572          1  1
2         1 0.56367 0.17262  0  1
3         1 0.56548 0.22739  1  0
4         0          0      0  0
```

wir so interpretiert:

Das saturierte Modell ist signifikant. Aber auch die einfacheren Modelle 2 und 3 sind mit den Daten verträglich. Da diese beiden Modelle gleich komplex sind, entscheiden wir uns für Modell 3, da es den größeren P-Wert besitzt.

## 5.2.4 Aufgaben

Die folgenden Aufgaben sollen mit quantitativen Variablen der Datensätze **statlab** und **kreditscoring** und zwei Prädiktoren durchgeführt werden.

(5.36) **Aufgabe:** Berechnen Sie die Regressionskoeffizienten mit Hilfe der R-Funktion **lm**.

(5.37) **Aufgabe:** Berechnen Sie die Regressionskoeffizienten mit Hilfe der expliziten Formeln.

(5.38) **Aufgabe:** Stellen Sie die Designmatrix auf und berechnen Sie die Regressionskoeffizienten mit Hilfe der Normalgleichungen.

(5.39) **Aufgabe:** Berechnen Sie die Abweichungsquadratsummen für die ANOVA-Tabelle ohne die Funktionen **summary** und **anova** zu verwenden.

(5.40) **Aufgabe:** Stellen Sie ANOVA-Tabelle (Funktion **anova**) auf und interpretieren Sie sie hinsichtlich des Signifikanzproblems. Stellen Sie die Varianzanalyse schematisch als Diagramm dar. Machen Sie das für alle möglichen Modelle.

(5.41) **Aufgabe:** Führen Sie eine Modellselektion durch.

### (5.42) TESTFRAGEN

1. Was ist ein zweifaches lineares Regressionsmodell ? Erklären Sie die Begriffe Regressionsparameter und Residuen !
2. Welche Erklärungsmodelle können bei einem zweifachen Regressionsmodell geprüft werden ?
3. Wie definiert man den multiplen Korrelationskoeffizienten beim zweifachen Regressionsmodell ?
4. Wie hängt der multiple Korrelationskoeffizient mit dem Erklärungsbeitrag zusammen ?
5. Was ist das Bestimmtheitsmass ? Wie hängt der multiple Korrelationskoeffizient mit dem Bestimmtheitsmass zusammen ?
6. Was versteht man unter dem saturierten Modell ?
7. Wie wird eine Modellselektion durchgeführt ?

## 5.3 Ein qualitativer Prädiktor

### 5.3.1 Die Fragestellung

Wir beschäftigen uns in diesem Kapitel mit statistischen Erklärungsmodellen, bei denen eine quantitative Responsevariable durch eine oder mehrere Prädiktorvariable erklärt werden sollen. Bis jetzt haben wir nur **Regressionsmodelle** betrachtet, dh. solche Erklärungsmodelle, bei denen alle Prädiktorvariablen vom quantitativen Typ sind. Außerdem haben wir uns auf **lineare** Regressionsmodelle beschränkt, dh. auf Modelle, bei denen die Regressionsfunktion eine lineare Funktion ist.

Nun wenden wir uns der Frage zu, wie man Erklärungsmodelle behandeln kann, bei denen die Prädiktoren auch qualitative Variable sein können. Zunächst wird uns längere Zeit der Fall beschäftigen, bei dem ein einziger Prädiktor auftritt, der außerdem eine qualitative Variable ist.

### 5.3.2 Prädiktoren mit zwei Kategorien

#### Modell und Modellvergleich

---

Ein qualitativer Prädiktor mit zwei Kategorien ist einfach eine Aufteilung der Untersuchungsobjekte in zwei Gruppen. Der Einfluß eines solchen Prädiktors auf die Responsevariable kann sich nur so äußern, dass die Responsevariable in jeder der beiden Gruppen eine andere Verteilung besitzt. In unserem Fall interessiert uns nur der Einfluss auf den Erwartungswert bzw. den Mittelwert.

Es sei also  $X$  ein Prädiktor mit den Kategorien  $a_1$  und  $a_2$ . Unser Modell besteht darin, dass die Responsevariable  $Y$  auf den beiden Gruppen  $G_1 = \{i : x_i = a_1\}$  und  $G_2 = \{i : x_i = a_2\}$  unterschiedliche Werte  $\mu_1$  und  $\mu_2$  annimmt (bis auf einen zufälligen Rest  $\epsilon_i$ ):

$$y_i = \begin{cases} \mu_1 + \epsilon_i & \text{wenn } i \in G_1 \\ \mu_2 + \epsilon_i & \text{wenn } i \in G_2 \end{cases}$$

Das ist natürlich ein Regressionsmodell. Um das rasch zu sehen, nehmen wir der Einfachheit halber an, dass  $G_1 = \{1, 2, \dots, n_1\}$  und  $G_2 = \{n_1 + 1, n_1 + 2, \dots, n\}$ :

$$\begin{aligned} y_1 &= 1 \cdot \mu_1 + 0 \cdot \mu_2 + \epsilon_1 \\ y_2 &= 1 \cdot \mu_1 + 0 \cdot \mu_2 + \epsilon_2 \\ &\vdots \\ y_{n_1} &= 1 \cdot \mu_1 + 0 \cdot \mu_2 + \epsilon_{n_1} \\ y_{n_1+1} &= 0 \cdot \mu_1 + 1 \cdot \mu_2 + \epsilon_{n_1+1} \\ &\vdots \\ y_n &= 0 \cdot \mu_1 + 1 \cdot \mu_2 + \epsilon_n \end{aligned}$$

Das Modell besitzt also die Designmatrix

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

Wir können dieses Modell mit der LSQ-Methode sehr leicht anpassen. Es ist bereits ohne Berechnungen klar, was als Parameterschätzungen herauskommen wird. Natürlich die Gruppenmittelwerte:

$$\hat{\mu}_1 = \bar{y}_{G_1} = \frac{y_1 + y_2 + \dots + y_{n_1}}{n_1} \quad \text{und} \quad \hat{\mu}_2 = \bar{y}_{G_2} = \frac{y_{n_1+1} + y_{n_1+2} + \dots + y_n}{n_2}$$

Dies liegt daran, dass innerhalb jeder Gruppe der Mittelwert jene Konstante ist, die nach dem LSQ-Prinzip optimal ist.

(5.43) BEISPIEL:

Wir stellen die Frage, wie gut im Datensatz **statlab** die Variable CBSEX die Variable CTPEA erklärt. In anderen Worten: Hat das Geschlecht Einfluß auf die Intelligenz ?

Wir führen die Modellanpassung auf zwei verschiedene Arten durch.

1. Gruppenmittelwerte:

```
> table(CBSEX)
CBSEX
  0    1
648 648
> tapply(CTPEA, CBSEX, mean)
      0      1
77.21142 80.96142
```

Wir sehen aus diesen Tabellen, dass die beiden Gruppen gleich groß sind, und dass in der zweiten Gruppe der Mittelwert höher ist als in der ersten Gruppe (Mädchen sind klüger !).

2. LSQ-Anpassung:

Wir überzeugen uns nun, dass die LSQ-Anpassung zum gleichen Ergebnis führt. Wir stellen die Designmatrix auf:

```
> X=cbind(CBSEX==0, CBSEX==1)
> X[1:3, ]
      [,1] [,2]
[1,] TRUE FALSE
[2,] TRUE FALSE
[3,] TRUE FALSE
> X[1294:1296, ]
      [,1] [,2]
[1,] FALSE TRUE
[2,] FALSE TRUE
[3,] FALSE TRUE
```

und lösen die Normalgleichungen:

```
> mu=solve(t(X)**X)**t(X)**CTPEA
> mu

      [,1]
[1,] 77.21142
[2,] 80.96142
```

Als Resultat erhalten wir die Gruppenmittelwerte.

Die Anpassung des 2-Gruppen-Modells alleine gibt uns noch keine Auskunft darüber, wie groß bzw. signifikant der Einfluss des Prädiktors auf die Responsevariable ist.

Wie kann man den Einfluss des Prädiktors auf die Responsevariable messen? Dafür gibt es zwei Möglichkeiten, die aber, wie wir sehen werden, eng miteinander zusammenhängen.

1. Wenn die beiden Gruppenmittelwerte gleich groß sind, dann hat der Prädiktor offenbar keinen Einfluss. Ein Einfluss des Prädiktors zeigt sich also am Unterschied der Gruppenmittelwerte.
2. Wenn das 2-Gruppen-Modell einen höheren Erklärungswert besitzt als das Konstantenmodell, dann ist das auch ein Hinweis auf einen Einfluss des Prädiktors auf die Responsevariable.

Die erste der beiden Möglichkeiten ist intuitiv näherliegend als die zweite. Deshalb spielen in der Praxis sogenannte Zweistichprobentests eine sehr große Rolle, bei denen es eben um den Einfluß eines dichotomen Prädiktors auf eine andere Variable (die Responsevariable) geht. Wenn allerdings der Prädiktor mehr als zwei Kategorien besitzt, dann tritt die zweite der beiden Möglichkeiten in den Vordergrund. Es ist daher wichtig, schon im einfachsten Fall (mit zwei Kategorien) die Funktionsweise zu verstehen.

(5.44) BEISPIEL:

Wir vergleichen das Modell  $CTPEA \sim CBSEX$  mit dem Modell  $CTPEA \sim 1$ .

Wir passen zunächst das Konstantenmodell an:

```
> fm.0=lm(CTPEA~1)
```

Anschließend passen wir das 2-Gruppenmodell an. Wir machen das jetzt mit der Funktion **lm**:

```
> X1=(CBSEX==0)
> X2=(CBSEX==1)
> fm.1=lm(CTPEA~0+X1+X2)
> summary(fm.1)
```

Call:

```
lm(formula = CTPEA ~ 0 + X1 + X2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-27.9614	-6.9614	0.7886	6.0386	49.0386

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t )
X1FALSE	80.9614	0.4087	198.1	<2e-16 ***
X1TRUE	77.2114	0.4087	188.9	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.4 on 1294 degrees of freedom

Multiple R-Squared: 0.983, Adjusted R-squared: 0.983

F-statistic: 3.747e+04 on 2 and 1294 DF, p-value: < 2.2e-16

Wir sehen, dass die Parameterschätzungen genau die Gruppenmittelwerte sind, wie wir ja erwarten. Die übrigen Teile der Tabelle sind uninteressant, da sie die Hypothesen abtesten, dass die Gruppenmittelwerte gleich Null sind (was selbst heutzutage eine unrealistische Hypothese ist).

Nun folgt der Vergleich der Modelle:

```
> anova(fm.0, fm.1)
```

Analysis of Variance Table

Model 1: CTPEA ~ 1

Model 2: CTPEA ~ 0 + X1 + X2

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	144596				
2	1294	140040	1	4556	42.101	1.231e-10 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Wir sehen, dass der Erklärungswert des 2-Gruppen-Modells hochsignifikant ist. Daraus schließen wir, dass CBSEX einen Einfluss auf CTPEA hat.

## Parametrisierung durch Haupteffekte

Grundsätzlich haben wir im vorangehenden Abschnitt alles gesagt, was nötig ist, um die Fragestellung zu beantworten. Allerdings gibt es elegantere und informativere Möglichkeiten, das Problem anzugehen. Es geht uns ja eigentlich um die Unterschiede der Gruppenmittelwerte.

Unser 2-Gruppen-Modell lautet in seiner ursprünglichen Form

$$y_i = \begin{cases} \mu_1 + \epsilon_i & \text{wenn } i \in G_1 \\ \mu_2 + \epsilon_i & \text{wenn } i \in G_2 \end{cases}$$

Dabei sind  $\mu_1$  und  $\mu_2$  die theoretischen Gruppenmittelwerte (Erwartungswerte). Wir bezeichnen mit

$$\bar{\mu} := \frac{\mu_1 + \mu_2}{2}$$

den Durchschnittswert. Wenn zwischen den Gruppen  $G_1$  und  $G_2$  kein Unterschied besteht, dann gilt

$$\mu_1 = \mu_2 = \bar{\mu} \quad \implies \quad \mu_1 - \bar{\mu} = \mu_2 - \bar{\mu} = 0$$

Jeder Unterschied zwischen  $\mu_1$  und  $\mu_2$  führt dazu, dass auch die Differenzen zum Mittelwert von Null verschieden werden:

$$\mu_1 \neq \mu_2 \quad \implies \quad \mu_1 - \bar{\mu} = -(\mu_2 - \bar{\mu}) = \frac{\mu_1 - \mu_2}{2} \neq 0$$

(5.45) DEFINITION:

Man nennt die Unterschiede zwischen den Gruppenmittelwerten  $\mu_1$  und  $\mu_2$  einerseits und ihrem Durchschnittswert andererseits die Haupteffekte des Prädiktors  $X$ :  $\mu_1 - \bar{\mu}$ ,  $\mu_2 - \bar{\mu}$

Es besteht nun die Möglichkeit, das Regressionsmodell  $Y \sim X$  so zu parametrisieren, dass einer der beiden Haupteffekte zugleich Regressionsparameter des Modells ist.

Wir definieren nun das Regressionsmodell  $Y \sim X$  so, dass der erste Haupteffekt Regressionsparameter ist. Zu diesem Zweck definieren wir

$$\begin{aligned} \beta_0 = \bar{\mu} &= \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 \\ \beta_1 = \mu_1 - \bar{\mu} &= \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 \end{aligned}$$

(Wir brauchen nur zwei Regressionsparameter, weil es auch nur zwei Gruppenmittelwerte gibt.) Aus diesem Gleichungssystem ergibt sich, wie wir im ursprünglichen Modell die Parameter  $\mu_1$  und  $\mu_2$  zu ersetzen haben. Wir lösen das Gleichungssystem

```
> C=cbind(c(0.5,0.5),c(0.5,-0.5))
> C
      [,1] [,2]
[1,]  0.5  0.5
[2,]  0.5 -0.5
> solve(C)
      [,1] [,2]
[1,]    1    1
[2,]    1   -1
```

und erhalten

$$\begin{aligned} \mu_1 &= \beta_0 + \beta_1 \\ \mu_2 &= \beta_0 - \beta_1 \end{aligned}$$

Man nennt die Matrix  $C$  die Kontrastmatrix des Modells. Aus der Inversen der Kontrastmatrix gewinnen wir die Designmatrix unseres neuen Regressionsmodells:

$$y_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{wenn } i \in G_1 \\ \beta_0 - \beta_1 + \epsilon_i & \text{wenn } i \in G_2 \end{cases}$$

Die Designmatrix lautet also:

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{pmatrix}$$

Gegenüber unserem ursprünglichen 2-Gruppen-Modell hat dieses Modell den Vorteil, dass der Regressionsparameter  $\beta_1$  die Information über den Unterschied zwischen den Gruppen angibt.

(5.46) BEISPIEL:

Wir parametrisieren das Modell  $CTPEA \sim CBSEX$  mit Haupteffekten. Das machen wir von Anfang an, obwohl wir im Grunde schon wissen, wie die Designmatrix aussehen muss.

Zunächst benötigen wir Hilfsvektoren:

```
> e=c(1,1)
> e1=c(1,0)
> e2=c(0,1)
```

Die Kontrastmatrix hat die Form

```
> C=rbind(e/2,e1-e/2)
> C
      [,1] [,2]
[1,]  0.5  0.5
[2,]  0.5 -0.5
```

Durch Invertieren erhalten wir die Koeffizienten der Designmatrix:

```
> coef.design=solve(C)
> coef.design
      [,1] [,2]
[1,]    1    1
[2,]    1   -1
```

Die erste Spalte der Designmatrix ist konstant =1. Um die zweite Spalte zu erhalten, müssen wir die Variable CBSEX mit den entsprechenden Koeffizienten der Designmatrix verschlüsseln:

```
> dummy=coef.design[factor(CBSEX),2]
> dummy[1:5]
[1] 1 1 1 1 1
> dummy[1292:1296]
[1] -1 -1 -1 -1 -1
```

Dieser Vorgang ist so zu erklären: Die Funktion **factor** kodiert die Werte von CBSEX mit Nummern 1,2,... Diese Nummern setzen wir dann als Indizes in die zweite Spalte der Koeffizientenmatrix ein.

Unser Modell hat nun die Variablen 1 (Konstante) und **dummy** (die Kodierung des ersten Haupteffekts). Wir passen das Modell an und sehen uns das Ergebnis an:

```
> fm=lm(CTPEA~dummy)
> summary(fm)
```

Call:

```
lm(formula = CTPEA ~ dummy)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-27.9614	-6.9614	0.7886	6.0386	49.0386

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
--	----------	------------	---------	----------

```
(Intercept) 79.086      0.289 273.681 < 2e-16 ***
dummy       -1.875      0.289  -6.489 1.23e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 10.4 on 1294 degrees of freedom
Multiple R-Squared:  0.03151,    Adjusted R-squared:  0.03076
F-statistic: 42.1 on 1 and 1294 DF,  p-value: 1.231e-10
```

Die Parameterschätzungen haben genau die Werte, die wir erwarten. Die Signifikanz des Parameters von **dummy** ist nichts anderes als die Signifikanz des ersten Haupteffekts, also des Gruppenunterschieds.

### 5.3.3 Prädiktoren mit mehr als zwei Kategorien

#### Parametrisierung durch Haupteffekte

Wir führen den Vorgang an einem Beispiel durch. Unsere Responsevariable sei CTPEA und die Prädiktorvariable sei MTE. MTE ist eine qualitative Variable mit 5 Ausprägungen.

Es gibt also 5 Gruppen von Familien, die den 5 Ausprägungen der Variablen MTE entsprechen. In jeder Gruppe hat CTPEA einen Mittelwert. Wenn wir die theoretischen Mittelwerte (Erwartungswerte) mit  $\mu_k$ ,  $k = 1, 2, 3, 4, 5$ , bezeichnen, dann kommen wir zum Modell

$$y_i = \begin{cases} \mu_1 + \epsilon_i & \text{wenn } i \in G_1 \\ \mu_2 + \epsilon_i & \text{wenn } i \in G_2 \\ \vdots & \\ \mu_5 + \epsilon_i & \text{wenn } i \in G_5 \end{cases}$$

Wir wollen aber nicht mit den Gruppenmittelwerten parametrisieren, sondern mit den Haupteffekten

$$\mu_1 - \bar{\mu}, \mu_2 - \bar{\mu}, \dots, \mu_5 - \bar{\mu}$$

Wir machen daher den Ansatz

$$\begin{aligned} \beta_0 &= \bar{\mu} \\ \beta_1 &= \mu_1 - \bar{\mu} \\ \beta_2 &= \mu_2 - \bar{\mu} \\ \beta_3 &= \mu_3 - \bar{\mu} \\ \beta_4 &= \mu_4 - \bar{\mu} \end{aligned}$$

Wir brauchen nur 4 Haupteffekte, weil wir insgesamt nur 5 Regressionsparameter benötigen. Dieses Gleichungssystem hat als Matrix die Kontrastmatrix:

```
> e=rep(1,5)
> e1=c(1,0,0,0,0)
> e2=c(0,1,0,0,0)
> e3=c(0,0,1,0,0)
> e4=c(0,0,0,1,0)
> C=rbind(e/5,e1-e/5,e2-e/5,e3-e/5,e4-e/5)
> C
      [,1] [,2] [,3] [,4] [,5]
```

```
[1,] 0.2 0.2 0.2 0.2 0.2
[2,] 0.8 -0.2 -0.2 -0.2 -0.2
[3,] -0.2 0.8 -0.2 -0.2 -0.2
[4,] -0.2 -0.2 0.8 -0.2 -0.2
[5,] -0.2 -0.2 -0.2 0.8 -0.2
```

Wir rechnen daraus die Koeffizienten der Designmatrix aus:

```
> coef.design=round(solve(C),5)
> coef.design
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    1    0    0    0
[2,]    1    0    1    0    0
[3,]    1    0    0    1    0
[4,]    1    0    0    0    1
[5,]    1   -1   -1   -1   -1
```

Diese Matrix gibt uns an, wie wir die Prädiktorvariable MTE verschlüsseln müssen: Durch 4 Dummyvariable, die den 4 letzten Spalten der Koeffizientenmatrix entsprechen. Wir führen nun die Verschlüsselung durch und passen das Modell an:

```
> dummy=coef.design[factor(MTE),2:5]
> fm=lm(CTPEA~dummy)
> summary(fm)
```

Call:

```
lm(formula = CTPEA ~ dummy)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-29.9313  -6.7490   0.2180   6.0687  50.1849
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  76.81720    0.46933  163.675 < 2e-16 ***
dummy1       -7.70181    1.57236  -4.898 1.09e-06 ***
dummy2       -3.34195    0.89443  -3.736 0.000195 ***
dummy3       -0.06821    0.58206  -0.117 0.906729
dummy4        2.99791    0.61053   4.910 1.03e-06 ***
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 9.879 on 1291 degrees of freedom

Multiple R-Squared: 0.1287, Adjusted R-squared: 0.126

F-statistic: 47.67 on 4 and 1291 DF, p-value: < 2.2e-16

Bevor wir uns mit der Interpretation dieser Ergebnisse auseinandersetzen, sehen wir uns an, wie R den Vorgang automatisiert.

Wenn R in einer Modellformel mit einem qualitativen Prädiktor konfrontiert wird (einem „Faktor“ in der Terminologie von R), dann verschlüsselt R diesen Prädiktor grundsätzlich in genau der gleichen Art und Weise, wie wir das gemacht haben. R speichert die konkreten Koeffizienten für die Verschlüsselung in einem Attribut mit der Bezeichnung **contrasts**.

Es gibt viele verschiedene Arten von Verschlüsselungen, die sich in den Koeffizienten unterscheiden. Die Voreinstellung von R lautet:

```
> contrasts(MTE)
  1 2 3 4
0 0 0 0 0
1 1 0 0 0
2 0 1 0 0
3 0 0 1 0
4 0 0 0 1
```

Diese Wahl der Koeffizienten für die Designmatrix kennen wir noch nicht und sie sagt uns auch nichts. (Wir lernen im nächsten Unterabschnitt, wie man feststellen kann, welche inhaltliche Interpretation diese Koeffizienten haben.)

Wir ändern nun die Voreinstellung so ab, dass unser Modell verwendet wird, bei dem die Regressionsparameter den ersten 4 Haupteffekten entsprechen. R liefert uns die benötigten Koeffizienten durch

```
> contr.sum(5)
  [,1] [,2] [,3] [,4]
1     1     0     0     0
2     0     1     0     0
3     0     0     1     0
4     0     0     0     1
5    -1    -1    -1    -1
```

Im R-Jargon würde man sagen: Das sind die Modellkoeffizienten der „Summenkontraste“. Um zu erreichen, dass R bei der Modellanpassung die Designmatrix mit den Summenkontrasten verschlüsselt, ändern wir das Kontrastattribut des Prädiktors:

```
> contrasts(MTE)=contr.sum(5)
> contrasts(MTE)
  [,1] [,2] [,3] [,4]
0     1     0     0     0
1     0     1     0     0
2     0     0     1     0
3     0     0     0     1
4    -1    -1    -1    -1
```

Nun können wir die Modellanpassung so durchführen, wie wir es gewohnt sind:

```
> fm=lm(CTPEA~MTE)
> summary(fm)
```

Call:

```
lm(formula = CTPEA ~ MTE)
```

Residuals:

Min	1Q	Median	3Q	Max
-29.9313	-6.7490	0.2180	6.0687	50.1849

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	76.81720	0.46933	163.675	< 2e-16 ***

```

MTE1      -7.70181      1.57236     -4.898 1.09e-06 ***
MTE2      -3.34195      0.89443     -3.736 0.000195 ***
MTE3      -0.06821      0.58206     -0.117 0.906729
MTE4       2.99791      0.61053      4.910 1.03e-06 ***

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 9.879 on 1291 degrees of freedom

Multiple R-Squared: 0.1287, Adjusted R-squared: 0.126

F-statistic: 47.67 on 4 and 1291 DF, p-value: < 2.2e-16

Als Ergebnis erhalten wir, dass lediglich der dritte Haupteffekt nicht signifikant ist. Wir überzeugen uns noch, dass die Parameterschätzungen tatsächlich mit den Haupteffekten übereinstimmen:

```

> m=tapply(CTPEA,MTE,mean)
> m
      0      1      2      3      4
69.11538 73.47525 76.74899 79.81510 84.93127
> mu.bar=mean(m)
> mu.bar
[1] 76.8172
> m-mu.bar
      0      1      2      3      4
-7.70181451 -3.34195160 -0.06821127  2.99790504  8.11407235

```

Die ersten beiden Gruppen haben signifikant unterdurchschnittliche Gruppenmittelwerte, die vierte Gruppe (Kode 3) hat einen signifikant überdurchschnittlichen Gruppenmittelwert.

Der Gruppenmittelwert der fünften Gruppe (Kode 4) ist auch auffällig groß. Allerdings haben wir keine Information über die Signifikanz, da sich der fünfte Haupteffekt nicht unter den Regressionsparametern befindet.

## Parametrisierung mit gewählten Kontrasten

---

Die Haupteffekte sind nur eine von vielen Möglichkeiten, die Regressionsparameter für einen qualitativen Prädiktor zu definieren. Wir erklären nun den allgemeinen Vorgang.

Gegeben sei ein qualitativer Prädiktor mit  $k$  Ausprägungen. Wenn ein Regressionsmodell mit dieser Prädiktorvariablen aufgestellt werden soll, dann muss dieses Modell  $k$  Parameter besitzen, denn grundsätzlich kann jeder Gruppenmittelwert  $\mu_1, \mu_2, \dots, \mu_k$  bei der Modellanpassung frei gewählt werden. Man sagt, das Modell besitzt die Dimension  $k$  oder (in alter statistischer Terminologie)  $k$  Freiheitsgrade.

Allerdings vergleichen wir ein solches Modell stets mit dem Konstantenmodell, das selbst einen Freiheitsgrad besitzt (die Dimension 1 hat). Es sind daher nur  $k - 1$  Parameter, die durch eine Prädiktorvariable mit  $k$  Ausprägungen zusätzlich zur Modellanpassung (zur Erklärung der Responsevariablen) zu Verfügung stehen.

Um die Parametrisierung des Regressionsmodells festzulegen, müssen wir also  $k - 1$  Parameter definieren.

---

(5.47) BEISPIEL:

Die Variable MTE ist eine ordinale Variable. Es liegt daher nahe, sich für die Unterschiede der Gruppenmittelwerte zwischen aufeinanderfolgenden Gruppen zu interessieren. Da MTE 5 Ausprägungen hat, gibt es 4 solche Gruppenunterschiede.

Wir definieren daher:

$$\begin{aligned}\beta_1 &= \mu_2 - \mu_1 \\ \beta_2 &= \mu_3 - \mu_2 \\ \beta_3 &= \mu_4 - \mu_3 \\ \beta_4 &= \mu_5 - \mu_4\end{aligned}$$

Für  $\beta_0$  nehmen wir wie bisher der Durchschnitt aller Gruppenmittelwerte.

Die Umsetzung in R sieht so aus. Zunächst bilden wir die Kontrastmatrix:

```
> e=rep(1,5)
> e1=c(1,0,0,0,0)
> e2=c(0,1,0,0,0)
> e3=c(0,0,1,0,0)
> e4=c(0,0,0,1,0)
> e5=c(0,0,0,0,1)
> C=rbind(e/5,e2-e1,e3-e2,e4-e3,e5-e4)
> C
      [,1] [,2] [,3] [,4] [,5]
[1,]  0.2  0.2  0.2  0.2  0.2
[2,] -1.0  1.0  0.0  0.0  0.0
[3,]  0.0 -1.0  1.0  0.0  0.0
[4,]  0.0  0.0 -1.0  1.0  0.0
[5,]  0.0  0.0  0.0 -1.0  1.0
```

Durch Invertieren erhalten wir die Koeffizienten der Designmatrix:

```
> ct=solve(C)
> ct
      [,1] [,2] [,3] [,4] [,5]
[1,]    1 -0.8 -0.6 -0.4 -0.2
[2,]    1  0.2 -0.6 -0.4 -0.2
[3,]    1  0.2  0.4 -0.4 -0.2
[4,]    1  0.2  0.4  0.6 -0.2
[5,]    1  0.2  0.4  0.6  0.8
```

Zur Parametrisierung benötigen wir nur die letzten 4 Spalten:

```
> contrasts(MTE)=ct[,2:5]
> contrasts(MTE)
      [,1] [,2] [,3] [,4]
0 -0.8 -0.6 -0.4 -0.2
1  0.2 -0.6 -0.4 -0.2
2  0.2  0.4 -0.4 -0.2
3  0.2  0.4  0.6 -0.2
4  0.2  0.4  0.6  0.8
```

Nun können wir die Modellanpassung durchführen:

```
> fm=lm(CTPEA~MTE)
> summary(fm)
```

Call:

```
lm(formula = CTPEA ~ MTE)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-29.9313	-6.7490	0.2180	6.0687	50.1849

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	76.8172	0.4693	163.675	< 2e-16	***
MTE1	4.3599	2.1725	2.007	0.04497	*
MTE2	3.2737	1.0788	3.035	0.00246	**
MTE3	3.0661	0.6721	4.562	5.54e-06	***
MTE4	5.1162	0.7678	6.664	3.95e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.879 on 1291 degrees of freedom

Multiple R-Squared: 0.1287, Adjusted R-squared: 0.126

F-statistic: 47.67 on 4 and 1291 DF, p-value: < 2.2e-16

Diese Art der Parametrisierung ist die richtige, wenn die Prädiktorvariable vom ordinalen Typ ist. Leider befinden sich die dafür nötigen Koeffizienten nicht in der Basisausstattung von R. Man muss die Library MASS dazuladen. Dann sind die Koeffizienten unter **contr.sdif** verfügbar.

```
> library(MASS)
> contrasts(MTE)=contr.sdif(5)
> fm=lm(CTPEA~MTE)
> summary(fm)
```

Call:

```
lm(formula = CTPEA ~ MTE)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-29.9313	-6.7490	0.2180	6.0687	50.1849

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	76.8172	0.4693	163.675	< 2e-16	***
MTE2-1	4.3599	2.1725	2.007	0.04497	*
MTE3-2	3.2737	1.0788	3.035	0.00246	**
MTE4-3	3.0661	0.6721	4.562	5.54e-06	***
MTE5-4	5.1162	0.7678	6.664	3.95e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.879 on 1291 degrees of freedom

Multiple R-Squared: 0.1287, Adjusted R-squared: 0.126

F-statistic: 47.67 on 4 and 1291 DF, p-value: < 2.2e-16

Wenn man diese Vorgangsweise wählt, dann zeigt die Beschriftung auch noch an, um welche Gruppenunterschiede es sich handelt. Wir können uns leicht davon überzeugen, dass

die Parameterschätzungen mit den Differenzen der Gruppenmittelwerte übereinstimmen.

Bei der Bildung der Koeffizientenmatrix durch Invertieren der Kontrastmatrix ist es bisher immer so gewesen, dass die Koeffizientenmatrix eine konstante erste Spalte besitzt, die aus 1-en besteht. Das war zwar unser Wunsch, aber warum es tatsächlich passiert ist, verstehen wir noch nicht. Die Sache ist aber ganz einfach. Die Kontrastmatrix muss immer folgende Bedingungen erfüllen:

- Die Summe der erste Zeile muss gleich 1 sein.
- Die Summen der übrigen Zeilen müssen alle gleich 0 sein.

Genau dann, wenn diese Bedingungen erfüllt sind (und die Kontrastmatrix invertierbar ist), hat die Koeffizientenmatrix eine erste Spalte, die aus 1-en besteht.

Wir können nun die Frage beantworten, welche Interpretation jene Koeffizienten haben, die von R als Voreinstellung zur Kodierung von qualitativen Prädiktoren verwendet werden.

Wir ergänzen die Koeffizientenmatrix mit einer ersten Spalte, die aus lauter 1-en besteht

```
> coef=cbind(rep(1,5),contr.treatment(5))
> coef
      2 3 4 5
1 1 0 0 0 0
2 1 1 0 0 0
3 1 0 1 0 0
4 1 0 0 1 0
5 1 0 0 0 1
```

und invertieren sie. Dadurch erhalten wir die Kontrastmatrix:

```
> solve(coef)
      1 2 3 4 5
      1 0 0 0 0
2 -1 1 0 0 0
3 -1 0 1 0 0
4 -1 0 0 1 0
5 -1 0 0 0 1
```

Daraus lesen wir die folgenden Interpretationen der Regressionsparameter ab:

$$\begin{aligned}\beta_0 &= \mu_1 \\ \beta_1 &= \mu_2 - \mu_1 \\ \beta_2 &= \mu_3 - \mu_1 \\ \beta_3 &= \mu_4 - \mu_1 \\ \beta_4 &= \mu_5 - \mu_1\end{aligned}$$

Eine solche Parametrisierung verwendet man dann, wenn die erste Gruppe eine Kontrollgruppe ist, die mit den übrigen Gruppen, deren Untersuchungsobjekte gewissen Behandlungen unterliegen, verglichen werden soll.

## 5.3.4 Aufgaben

(5.48) **Aufgabe:** Untersuchen Sie den Einfluss von FTE auf FIT.

(5.49) **Aufgabe:** Untersuchen Sie den Einfluss von FTO auf FIT.

(5.50) **Aufgabe:** Untersuchen Sie den Einfluss von CBB auf CBWGT.

(5.51) **Aufgabe:** Untersuchen Sie den Einfluss von CTL auf CTPEA.

(5.52) TESTFRAGEN

1. Durch wieviele Dummyvariable wird eine qualitative Prädiktorvariable in einem linearen Regressionsmodell kodiert ?
2. Wie erreicht man, dass die Regressionsparameter der Dummyvariablen eine bestimmte inhaltliche Bedeutung haben ?
3. Was ist der Unterschied zwischen der Kontrastmatrix und der Designmatrix eines qualitativen Prädiktors ?
4. Wie kodiert man ordinale Prädiktoren ? Wie interpretiert man die Regressionsparameter ?
5. Was Treatment-Kontraste ? Wie interpretiert man die Regressionsparameter ?
6. Was sind die Haupteffekte eines qualitativen Prädiktors ?

## 5.4 Additive Modelle und Wechselwirkungen

Wir haben bisher nur solche Regressionsmodelle betrachtet, bei denen jede Erklärungsvariable durch einen festen Regressionsparameter auf die Responsevariable einwirkt. Wenn es mehrere Erklärungsvariable gibt, dann kann es aber auch sein, dass der Regressionsparameter eines Prädiktors von einem anderen Prädiktor abhängt. Man spricht dann von **Wechselwirkung** der Prädiktoren.

Grundsätzlich unterscheidet man

**Additive Modelle:** Die Regressionsparameter der Prädiktoren hängen nicht von anderen Prädiktoren ab.

**Modelle mit Wechselwirkungen:** Die Regressionsparameter der Prädiktoren hängen von anderen Prädiktoren ab.

Wir werden nun für Probleme mit zwei Prädiktoren Regressionsmodelle mit Wechselwirkungen untersuchen. Dabei müssen wir drei Fälle unterscheiden:

- Beide Prädiktoren sind quantitative Variable.
- Ein Prädiktor ist quantitativ und ein Prädiktor ist qualitativ.
- Beide Prädiktoren sind qualitative Variable.

### 5.4.1 Zwei quantitative Prädiktoren

Gegeben sei eine Responsevariable  $Y$  und zwei quantitative Prädiktorvariable  $X_1$  und  $X_2$ . Ein zweifaches Regressionsmodell der Form

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

ist ein additives Modell. Das liegt daran, dass die Regressionskoeffizienten der einzelnen Prädiktoren nicht vom jeweils anderen Prädiktor abhängen.

Ein Modell mit Wechselwirkungen entsteht nun dadurch, dass wir zulassen, dass die Regressionskoeffizienten eines Prädiktors vom jeweils anderen Prädiktor abhängen. Wir modellieren diese Abhängigkeit durch eine lineare Funktion, also

$$\begin{aligned}\beta_1 &= \gamma_0 + \gamma_1 x_{2i} \\ \beta_2 &= \delta_0 + \delta_1 x_{1i}\end{aligned}$$

Wir setzen nun in die Modellgleichung ein und fassen zusammen:

$$\begin{aligned}y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \\ &= \beta_0 + (\gamma_0 + \gamma_1 x_{2i}) x_{i1} + (\delta_0 + \delta_1 x_{1i}) x_{i2} + \epsilon_i \\ &= \beta_0 + \gamma_0 x_{i1} + \delta_0 x_{i2} + (\gamma_1 + \delta_1) x_{i1} x_{i2} + \epsilon_i\end{aligned}$$

Es entsteht also eine Regressionsmodell mit drei Prädiktoren  $X_1$ ,  $X_2$  und  $X_1 \cdot X_2$ . Der dritte Prädiktor ist rechnerisch das Produkt der ersten beiden Prädiktoren und steht inhaltlich für den Effekt der Wechselwirkung.

Der allgemeine Ansatz für ein Regressionsmodell mit Wechselwirkung sieht also so aus:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$$

Für die Interpretation von Wechselwirkungen sind einige Grundsätze zu beachten: Wechselwirkungseffekte haben nur Sinn hinsichtlich ihres zusätzlichen Erklärungswerts zu den Haupteffekten. Da die Variable  $X_1 \cdot X_2$  im allgemeinen niemals unkorreliert ist zu  $X_1$  und  $X_2$ , saugt sie stets Erklärungswert von den Variablen  $X_1$  und  $X_2$  ab. Es ist daher aus den Parameterschätzungen im Wechselwirkungsmodell nicht ablesbar, ob die Parameter der Haupteffekte signifikant sind. Die Beurteilung der Signifikanz der Haupteffekte muss stets im kleineren additiven Modell erfolgen.

## (5.53) BEISPIEL:

Wir untersuchen im Datensatz **statlab** die Abhängigkeit der Variablen CTPEA von MBAG und FBAG.

```
> fm.add=lm(CTPEA~MBAG+FBAG)
> fm.ww=lm(CTPEA~MBAG+FBAG+MBAG:FBAG)
> anova(fm.add, fm.ww)
Analysis of Variance Table

Model 1: CTPEA ~ MBAG + FBAG
Model 2: CTPEA ~ MBAG + FBAG + MBAG:FBAG
  Res.Df    RSS    Df Sum of Sq      F Pr(>F)
1    1293 141941
2    1292 141694     1      247 2.2505 0.1338
```

Es gibt offenbar keine signifikante Wechselwirkung.

## (5.54) BEISPIEL:

Wir untersuchen nun im Datensatz **statlab** die Abhängigkeit der Variablen CTRA von MBAG und FBAG.

```
> fm.add=lm(CTRA~MBAG+FBAG)
> fm.ww=lm(CTRA~MBAG+FBAG+MBAG:FBAG)
> anova(fm.add, fm.ww)
Analysis of Variance Table

Model 1: CTRA ~ MBAG + FBAG
Model 2: CTRA ~ MBAG + FBAG + MBAG:FBAG
  Res.Df    RSS    Df Sum of Sq      F    Pr(>F)
1    1293 128728
2    1292 127319     1      1409 14.302 0.0001628 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Es besteht also eine signifikante Wechselwirkung. Aus den Parameterschätzungen

```
> summary(lm(CTRA~MBAG+FBAG+MBAG:FBAG))
```

Call:

```
lm(formula = CTRA ~ MBAG + FBAG + MBAG:FBAG)
```

Residuals:

Min	1Q	Median	3Q	Max
-23.2817	-7.5153	0.3790	7.8662	25.2852

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.443531	6.042079	1.232	0.218192
MBAG	0.847254	0.223660	3.788	0.000159 ***
FBAG	0.683101	0.198357	3.444	0.000592 ***
MBAG:FBAG	-0.023782	0.006288	-3.782	0.000163 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.927 on 1292 degrees of freedom

Multiple R-Squared: 0.01178, Adjusted R-squared: 0.009489

F-statistic: 5.135 on 3 and 1292 DF, p-value: 0.001562

lesen wir ab, dass das saturierte Modell nicht vereinfacht werden kann. Der Effekt der Wechselwirkung ist negativ, dh. der zweite Prädiktor wirkt vermindernd auf den Regressionsparameter des ersten Prädiktors, und umgekehrt:

$$\begin{aligned}\hat{Y} &= 7.44 + 0.85X_1 + 0.68X_2 - 0.02X_1 \cdot X_2 \\ &= 7.44 + (0.85 - 0.02X_2) \cdot X_1 + 0.68X_2 \\ &= 7.44 + 0.85X_1 + (0.68 - 0.02X_1) \cdot X_2\end{aligned}$$

## 5.4.2 Ein qualitativer und ein quantitativer Prädiktor

Es sei  $Y$  eine Responsevariable und  $X_1, X_2$  seien Prädiktoren. Dabei sei  $X_1$  eine qualitative Variable und  $X_2$  eine quantitative Variable.

### Das additive Modell

Der qualitative Prädiktor definiert eine Gruppierung, und geht ins Modell ein durch die Gruppenmittelwerte. Der quantitative Prädiktor geht über seinen Regressionsparameter ins Modell ein:

$$y_i = \mu_k + \beta_2 x_{i2} + \epsilon_i \quad \text{wenn } i \in G_k$$

In Worten läßt sich dieses Modell so beschreiben:

- In jeder Gruppe  $G_k$  wird der Zusammenhang zwischen  $Y$  und  $X_2$  durch eine Regressionsgerade beschrieben.
- Die Slopes der Regressionsgeraden sind für alle Gruppen gleich. Lediglich die Intercepts können sich unterscheiden.

Die praktische Bedeutung solcher additiver Modelle liegt vor allem in der Fragestellung der **Kovarianzanalyse**: Dabei geht es darum, dass primär der Einfluss des qualitativen Faktors  $X_1$  auf  $Y$  untersucht werden soll, wobei aber **vor** der Berechnung des Erklärungsbeitrags

von  $X_1$  der Einfluss der einer quantitativen **Kovariate** (Störvariable) ausgeschaltet werden soll.

(5.55) BEISPIEL:

Es soll im Datensatz **statlab** die Abhängigkeit der Intelligenz vom Geschlecht untersucht werden.

Die beiden Intelligenztests CTPEA und CTRA messen unterschiedliche Dimensionen der Intelligenz. CTPEA misst sprachliche und kulturabhängige Leistungen, während CTRA logisch-abstrakte und kulturunabhängige Leistungen misst. Natürlich sind CTRA und CTPEA stark korreliert, denn die allgemeine intellektuelle Leistungsfähigkeit spielt bei beiden Tests eine große Rolle.

Wenn man nun feststellen will, wie jede einzelne der Dimensionen der Intelligenz vom Geschlecht abhängt, muss man vorher dafür sorgen, dass die allgemeine intellektuelle Leistungsfähigkeit ausgeschaltet wird. Das macht man durch Einführung einer Kovariate ins Modell.

Für die Responsevariable CTPEA verwenden wir CTRA als Kovariate. Sie eliminiert jene Variation der Intelligenz, die beide Variable betrifft.

```
> fm.1=lm(CTPEA~CTRA)
> fm.2=lm(CTPEA~CTRA+CBSEX)
> anova(fm.1, fm.2)
Analysis of Variance Table
```

```
Model 1: CTPEA ~ CTRA
```

```
Model 2: CTPEA ~ CTRA + CBSEX
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1294	108849				
2	1293	102551	1	6298	79.403	< 2.2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wir sehen, dass das Geschlecht einen hochsignifikanten Einfluss auf die sprachliche Intelligenz hat. Um die Richtung des Einflusses zu untersuchen, sehen wir uns die Parameterschätzungen an:

```
> contrasts(CBSEX)=contr.sdif(2)
> coef(lm(CTPEA~CTRA+CBSEX))
(Intercept)          CTRA      CBSEX2-1
 62.3587849    0.5404621    4.4172371
```

```
> boxplot((CTPEA-0.54*CTRA)~CBSEX)
```

Das heisst: Nach Elimination der allgemeinen intellektuellen Leistungsfähigkeit verbleibt eine Differenz der Gruppenmittelwerte von 4.4 zugunsten der Mädchen bei der sprachlichen Intelligenz.

Nun ist es hoch an der Zeit, die Ehre der Burschen zu retten.

Wir untersuchen die logische Intelligenz CTRA auf die gleiche Weise, indem wir die allgemeine intellektuelle Leistungsfähigkeit durch CTPEA eliminieren:

```
> fm.1=lm(CTRA~CTPEA)
```

```

> fm.2=lm(CTRA~CTPEA+CBSEX)
> anova(fm.1, fm.2)
Analysis of Variance Table

Model 1: CTRA ~ CTPEA
Model 2: CTRA ~ CTPEA + CBSEX
  Res.Df  RSS    Df Sum of Sq      F    Pr(>F)
1     1294 96985
2     1293 93985     1      3000 41.273 1.857e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Auch hier entdecken wir einen hochsignifikanten Einfluss des Geschlechts. Die Parameterschätzung

```

> coef(lm(CTRA~CTPEA+CBSEX))
(Intercept)      CTPEA    CBSEX2-1
-8.2224020    0.4953192   -3.0920148

```

```

> boxplot((CTRA-0.49*CTPEA)~CBSEX)

```

zeigt uns aber, dass hier die Burschen deutlich überlegen sind.

Diese Unterschiede der Geschlechter in den beiden Dimensionen der Intelligenz sind wesentlich stärker ausgeprägt als der Unterschied in der allgemeinen intellektuellen Leistungsfähigkeit. Wir nehmen als Mass dafür die Summe aus beiden Intelligenztests und erhalten

```

> summary(lm(I(CTPEA+CTRA)~CBSEX))

```

Call:

```
lm(formula = I(CTPEA + CTRA) ~ CBSEX)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-47.2948 -12.2948   0.2207  12.3418  74.7052

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 110.0370     0.4927 223.332 <2e-16 ***
CBSEX2-1     2.5154     0.9854   2.553  0.0108 *
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 17.74 on 1294 degrees of freedom
Multiple R-Squared:  0.00501,    Adjusted R-squared:  0.004242
F-statistic: 6.516 on 1 and 1294 DF,  p-value: 0.01080

```

Der Unterschied ist hier zwar zugunsten der Mädchen, aber er ist nur halb so gross und vor allem wesentlich weniger signifikant. Wenn man bedenkt, dass 10-jährige Mädchen meist biologisch wesentlich reifer sind als 10-jährige Jungen, dann relativiert sich Rolle des Geschlechts als Erklärungsursache noch mehr.

---

Im vorangehenden Beispiel haben wir den Regressionskoeffizienten der Dummyvariablen von CBSEX als Mittelwertunterschied interpretiert. Dies erfordert eine Begründung.

Wenn wir als Kodierung verwenden

$$\begin{aligned}\mu_1 &= \beta_0 + \frac{1}{2}\beta_1 \\ \mu_2 &= \beta_0 - \frac{1}{2}\beta_1\end{aligned}$$

dann lautet das Modell

$$y_i = \begin{cases} \beta_0 + \frac{1}{2}\beta_1 + \beta_2 x_i + \epsilon_i & \text{wenn } i \in G_1 \\ \beta_0 - \frac{1}{2}\beta_1 + \beta_2 x_i + \epsilon_i & \text{wenn } i \in G_2 \end{cases}$$

Wenn wir Gruppenmittelwerte bilden, dann fallen bei optimalen Parameterwerten die Residuen weg, weil die Residuen zur Dummyvariablen des qualitativen Faktors orthogonal sind:

$$\begin{aligned}\bar{y}_{G_1} &= \hat{\beta}_0 + \frac{1}{2}\hat{\beta}_1 + \hat{\beta}_2 \bar{x}_{G_1} \\ \bar{y}_{G_2} &= \hat{\beta}_0 - \frac{1}{2}\hat{\beta}_1 + \hat{\beta}_2 \bar{x}_{G_2}\end{aligned}$$

Wir subtrahieren die Gleichungen

$$\bar{y}_{G_1} - \bar{y}_{G_2} = \hat{\beta}_1 + \hat{\beta}_2(\bar{x}_{G_1} - \bar{x}_{G_2})$$

und rechnen  $\hat{\beta}_1$  aus:

$$\hat{\beta}_1 = (\bar{y}_{G_1} - \hat{\beta}_2 \bar{x}_{G_1}) - (\bar{y}_{G_2} - \hat{\beta}_2 \bar{x}_{G_2})$$

Der Regressionsparameter ist also identisch mit dem Unterschied der Gruppenmittelwerte der bereinigten Responsevariablen.

(5.56) BEISPIEL:

Hängt das Alter des Ehemanns vom Bildungsstand der Frau ab ?

Wenn wir im Datensatz **statlab** die Abhängigkeit der Responsevariablen FBAG von der Variablen MTE untersuchen, dann erhalten wir:

```
> anova(lm(FBAG~1), lm(FBAG~MTE))
Analysis of Variance Table
```

```
Model 1: FBAG ~ 1
```

```
Model 2: FBAG ~ MTE
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	57456				
2	1291	55366	4	2090	12.182	1.005e-09 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wenn wir aber die Kovariate MBAG hinzufügen, dann ergibt sich ein anderes Bild:

```
> anova(lm(FBAG~MBAG), lm(FBAG~MBAG+MTE))
Analysis of Variance Table
```

```
Model 1: FBAG ~ MBAG
```

```
Model 2: FBAG ~ MBAG + MTE
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1294	18170.6				

```
2 1290 18048.8 4 121.9 2.1775 0.0694 .
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Das Modell mit Wechselwirkung

Der qualitative Prädiktor definiert wieder eine Gruppierung, und geht ins Modell ein durch die Gruppenmittelwerte. Der quantitative Prädiktor geht über die Regressionsparameter ins Modell ein, wobei aber nun jede Gruppe einen anderen Regressionsparameter haben kann:

$$y_i = \mu_k + \delta_k x_{i2} + \epsilon_i \quad \text{wenn } i \in G_k$$

In Worten läßt sich dieses Modell so beschreiben:

- In jeder Gruppe  $G_k$  wird der Zusammenhang zwischen  $Y$  und  $X_2$  durch eine Regressionsgerade beschrieben.
- Sowohl die Slopes der Regressionsgeraden als auch die Intercepts können sich zwischen den Gruppen unterscheiden.

Wenn man den Erklärungswert dieses Modells relativ zum additiven Modell untersucht, dann untersucht man die Frage, ob die Regressionsparameter in den einzelnen Gruppen voneinander verschieden sind.

Sehen wir uns einige Beispiele zunächst graphisch an.

(5.57) BEISPIEL:

Man kann die Streudiagramme und Regressionsgeraden von Datengruppen mit der Funktion **coplot** graphisch vergleichen. Die Panelfunktion **scatter2** stellen wir in der Library **LinAlg.r** zu Verfügung.

```
> coplot(CTPEA~MBAG|MTE, show.given=FALSE, panel=scatter2)
> coplot(CTRA~MBAG|MTE, show.given=FALSE, panel=scatter2)
> coplot(CTPEA~CTRA|CBSEX, show.given=FALSE, panel=scatter2)
> coplot(FIT~I(FBAG~MBAG)|MTE, show.given=FALSE, panel=scatter2)
> coplot(FIT~I(FBAG~MBAG)|FTE, show.given=FALSE, panel=scatter2)
> coplot(FIT~I(FBAG~MBAG)|CBSEX, show.given=FALSE, panel=scatter2)
> coplot(FBAG~MBAG|MTE, show.given=FALSE, panel=scatter2)
> coplot(FBAG~MBAG|FTE, show.given=FALSE, panel=scatter2)
> coplot(MTWGT~MTHGHT|MTE, show.given=FALSE, panel=scatter2)
> coplot(CTWGT~CTHGHT|MTE, show.given=FALSE, panel=scatter2)
```

Wenn wir ein Modell mit Wechselwirkung anpassen, dann verwenden wir natürlich Dummyvariable. Sehen wir uns den Fall eines qualitativen Faktors mit 2 Kategorien an. Wir ersetzen im Modell

$$y_i = \mu_k + \delta_k x_{i2} + \epsilon_i \quad \text{wenn } i \in G_k$$

die Parameter durch

$$\begin{aligned} \mu_1 &= \beta_0 + \frac{1}{2}\beta_1 \\ \mu_2 &= \beta_0 - \frac{1}{2}\beta_1 \end{aligned} \quad \implies \quad \beta_1 = \mu_1 - \mu_2$$

und

$$\begin{aligned} \delta_1 &= \beta_2 + \frac{1}{2}\beta_3 \\ \delta_2 &= \beta_2 - \frac{1}{2}\beta_3 \end{aligned} \quad \implies \quad \beta_3 = \delta_1 - \delta_2$$

Dadurch entsteht das Modell

$$\begin{aligned} y_i &= \mu_k + \delta_k x_{i2} + \epsilon_i \quad \text{wenn } i \in G_k \\ &= \begin{cases} \beta_0 + \frac{1}{2}\beta_1 + \beta_2 x_{i2} + \frac{1}{2}\beta_3 x_{i2} + \epsilon_i \\ \beta_0 - \frac{1}{2}\beta_1 + \beta_2 x_{i2} - \frac{1}{2}\beta_3 x_{i2} + \epsilon_i \end{cases} \end{aligned}$$

Der Regressionsparameter  $\beta_3$  steht also für die Wechselwirkung. Wir fassen zusammen:

- Der zu  $\beta_3$  gehörige Spaltenvektor der Designmatrix ist gleich dem Produkt aus der Variablen  $X_2$  und der Dummyvariablen von  $X_1$ .
- Die inhaltliche Bedeutung von  $\beta_3$  ist die Differenz der Slopes der Regressionsgeraden in den Gruppen.

(5.58) BEISPIEL:

Wir untersuchen die Abhängigkeit der Variablen FIT von FBAG-MBAG und von MTE. Aus der entsprechenden Graphik sehen wir deutliche Unterschiede in den Slopes der Regressionsgeraden. Wir wollen nun wissen, ob diese Unterschiede signifikant sind:

```
> X=I(FBAG-MBAG)
> fm.1=lm(FIT~X)
> fm.2=lm(FIT~X+MTE)
> fm.3=lm(FIT~X+MTE+MTE:X)
> anova(fm.1, fm.2, fm.3)
Analysis of Variance Table
```

```
Model 1: FIT ~ X
Model 2: FIT ~ X + MTE
Model 3: FIT ~ X + MTE + MTE:X
  Res.Df    RSS   Df Sum of Sq      F Pr(>F)
1     1294 5918115
2     1290 4963501     4    954615 62.1536 <2e-16 ***
3     1286 4937905     4     25596  1.6665 0.1553
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Die Unterschiede der Slopes sind also nicht signifikant.

(5.59) BEISPIEL:

Wir haben im vorangehenden Abschnitt festgestellt, dass es bei der Abhängigkeit der Variablen CTRA vom Alter der Eltern eine signifikante Wechselwirkung gibt, die außerdem ein negatives Vorzeichen hat.

Diese Ergebnis wollen wir uns nun nochmals näher ansehen. Zu diesem Zweck teilen wir die Mütter nach ihrem Alter in zwei Gruppen auf:

```
> dMBAG=cut(MBAG,2)
> table(dMBAG)
dMBAG
(13,29.5] (29.5,46]
      806      490
```

Die Graphik zeigt die negative Wechselwirkung sehr deutlich.

```
coplot(CTRA~FBAG|dMBAG,show.given=FALSE,panel=scatter2)
```

Wir überprüfen nun die Signifikanz:

```
> fm.0=lm(CTRA~1)
> fm.1=lm(CTRA~FBAG)
> fm.2=lm(CTRA~FBAG+dMBAG)
> fm.3=lm(CTRA~FBAG+dMBAG+FBAG:dMBAG)
> anova(fm.0, fm.1, fm.2, fm.3)
Analysis of Variance Table
```

```
Model 1: CTRA ~ 1
Model 2: CTRA ~ FBAG
Model 3: CTRA ~ FBAG + dMBAG
Model 4: CTRA ~ FBAG + dMBAG + FBAG:dMBAG
  Res.Df    RSS    Df Sum of Sq      F    Pr(>F)
1     1295 128837
2     1294 128781     1      56  0.5653 0.452256
3     1293 128758     1      23  0.2359 0.627263
4     1292 127688     1    1070 10.8273 0.001027 **
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Es ist eindrucksvoll zu sehen, dass eine einzelne Regressionsgerade oder zwei Regressionsgerade mit gleichem Slope überhaupt keinen nennenswerten Erklärungsbeitrag liefern. Erst nach Einführung des Wechselwirkungseffekts erkennt man den Erklärungsbeitrag der Variablen FBAG.

### 5.4.3 Zwei qualitative Variable

Lineare Erklärungsmodelle, bei denen eine quantitative Responsevariable durch zwei qualitative Variable erklärt werden soll, spielen eine überragende Rolle für viele Anwendungsbereiche. Wenn man von „zweifacher Varianzanalyse“ spricht, dann meint man solche Modelle.

#### Die Problemstellung

Es sei  $Y$  die Responsevariable und  $X_1$  und  $X_2$  seien die Prädiktorvariablen. Da die Prädiktorvariablen qualitativ sind, definieren sie jeweils eine Gruppeneinteilung. Bezeichnet  $G_1, G_2, \dots, G_p$  die durch  $X_1$  definierte Gruppierung und  $H_1, H_2, \dots, H_q$  die durch  $X_2$

definierte Gruppierung, dann haben wir es insgesamt mit einer Kreuztabelle zu tun:

	$H_1$	$H_2$	...	$H_q$
$G_1$	$G_1 \cap H_1$	$G_1 \cap H_2$	...	$G_1 \cap H_q$
$G_2$	$G_2 \cap H_1$	$G_2 \cap H_2$	...	$G_2 \cap H_q$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$G_p$	$G_p \cap H_1$	$G_p \cap H_2$	...	$G_p \cap H_q$

In jeder Zelle der Kreuztabelle können sich Beobachtungen befinden oder auch nicht.

Wir bezeichnen mit  $n_{k\ell}$  die Anzahl der Daten in Zelle  $G_k \cap H_\ell$ . Mit  $m_{k\ell}$  bezeichnen wir den Mittelwert der Responsevariablen  $Y$  in Zelle  $G_k \cap H_\ell$ . Wenn sich einer Zelle keine Daten befinden, dh. wenn  $n_{k\ell} = 0$  ist, dann ist der Mittelwert dort nicht definiert.

(5.60) BEISPIEL:

Eine Kreuztabellierung der Faktoren MTE und FTE ergibt:

```
> table(MTE, FTE)
      FTE
MTE  0   1   2   3   4
  0  12   5   5   3   1
  1  10  27  46  15   3
  2  18  66 200 146   64
  3   4  17  95 134  134
  4   0   3  20  38  230
```

Eine der Zellen enthält keine Daten. Die Tabelle der Zellmittelwerte enthält dort NA:

```
> round(tapply(FIT, list(MTE, FTE), mean), 2)
      0      1      2      3      4
0  87.92 107.80  77.00  85.33 100.00
1  88.70 115.22 121.80 116.40 171.33
2 121.44 121.47 133.59 140.72 173.06
3 154.00 133.06 145.61 151.13 170.78
4      NA 116.33 186.80 172.82 208.25
```

In Zusammenhang mit solchen Kreuztabellierungen spielen die folgenden Fragestellungen eine Rolle:

**Haupteffekt von  $X_1$ :** Liefert der Faktor  $X_1$  einen Beitrag zur Erklärung der Unterschiede zwischen den Zeilen ?

**Haupteffekt von  $X_2$ :** Liefert der Faktor  $X_2$  einen Beitrag zur Erklärung der Unterschiede zwischen den Spalten ?

**Wechselwirkung von  $X_1$  und  $X_2$ :** Können die Zellmittelwerte durch die beiden Haupteffekte alleine erklärt werden, oder liefert die Zugehörigkeit zu einer bestimmten Zelle einen zusätzlichen Erklärungsbeitrag ?

Wir wissen schon aus dem Abschnitt über die mehrfache Regression dass der Erklärungsbeitrag eines Prädiktor in der Regel davon abhängt, ob sich noch andere Prädiktoren im Modell befinden. Nur wenn die Prädiktoren unkorreliert sind, dann tritt diese Komplikation nicht auf.

Wenn die Werte der qualitativen Prädiktoren  $X_1$  und  $X_2$  durch zufällige Beobachtungen entstehen, dann sind die Dummyvariablen von  $X_1$  und  $X_2$  im allgemeinen durchaus korreliert, was daran liegt, dass die Variablen  $X_1$  und  $X_2$  stochastisch gekoppelt sind. bei der Beurteilung des Haupteffektes eines Prädiktors spielt es dann eine große Rolle, ob der zweite Prädiktor sich schon im Modell befindet oder nicht.

Haben wir es mit qualitativen Prädiktoren zu tun, dann können wir manchmal den Versuch planen. Das heisst, wir können vornherein festlegen, wieviele Daten sich in den einzelnen Zellen der Kreuztabellierung befinden sollen. Jede solche Festlegung nennt man einen **Versuchsplan**.

Um jene Schwierigkeiten mit der Interpretation der Haupteffekte zu vermeiden, die dadurch entstehen, dass die Dummyvariablen der Prädiktoren korreliert sind, ist es wünschenswert, die Zelhäufigkeiten so zu wählen, dass die Dummyvariablen unkorreliert sind. Wenn wir das erreichen, dann haben wir es mit einem **orthogonalen Versuchsplan** zu tun. Auf orthogonale Versuchspläne können wir nicht mehr eingehen.

## Additive Modelle

---

Das additive Modell bei zwei qualitativen Faktoren sieht folgendermaßen aus:

$$y_i = \mu_k + \nu_\ell + \epsilon_i \quad \text{wenn } i \in G_k \cap H_\ell$$

Dieses Modell ist nicht ganz so harmlos, wie es auf den ersten Blick aussieht. Die Modellparameter sind nämlich durch die Modellgleichung nicht eindeutig festgelegt. Man könnte ja zu allen  $\mu_k$  eine Zahl dazu addieren und die gleiche Zahl von allen  $\nu_\ell$  abziehen. Dadurch würde sich am Modell nichts ändern, sondern nur an der Parametrisierung. Da wir aber nur an der Unterschiedlichkeit der  $\mu_k$  und der  $\nu_\ell$  interessiert sind, spielt die Addition oder Subtraktion einer Konstanten keine Rolle.

---

(5.61) BEISPIEL:

Wir passen nun für die Responsevariable FIT das additive Modell MTE+FTE an. Zunächst machen wir das selbst, ohne die Funktion **lm** zu Hilfe zu nehmen.

Die Designmatrix  $X$  besteht aus den Indikatorvariablen der Faktoren MTE und FTE. Um diese Matrix aufzubauen, bilden wir zunächst die Einheitsmatrix

```
> eye=diag(rep(1,5))
> eye
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0    0
[2,]    0    1    0    0    0
[3,]    0    0    1    0    0
[4,]    0    0    0    1    0
[5,]    0    0    0    0    1
```

und konstruieren daraus die Designmatrix von MTE:

```
> X1=eye[factor(as.numeric(MTE)),]
> X1[1:5,]
      [,1] [,2] [,3] [,4] [,5]
```

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

Auf die gleiche Art und Weise konstruieren wir die Designmatrix von FTE:

```
> X2=eye[ factor(as.numeric(FTE)), ]
> X2[1:5, ]
      [,1] [,2] [,3] [,4] [,5]
[1,] 0 0 0 1 0
[2,] 0 1 0 0 0
[3,] 0 0 1 0 0
[4,] 0 0 0 1 0
[5,] 0 0 1 0 0
```

Die Designmatrix des additiven Modells ist dann

```
> X[1:5, ]
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 0 1 0 0 0 0 0 0 0 1 0
[2,] 0 0 0 1 0 0 0 1 0 0 0
[3,] 0 0 1 0 0 0 0 0 1 0 0
[4,] 0 0 1 0 0 0 0 0 0 1 0
[5,] 0 0 1 0 0 0 0 0 1 0 0
```

Optimale Schätzungen für unsere Parameter

$$\theta = (\mu_1, \mu_2, \dots, \mu_5, \nu_1, \nu_2, \dots, \nu_5)^t$$

erhalten wir durch Lösen der Normalgleichungen

$$(\mathbf{X}^t \mathbf{X}) \theta = \mathbf{X}^t \mathbf{y}$$

Auf Grund der Unbestimmtheit der Modellparameter ist aber die Matrix  $\mathbf{X}^t \mathbf{X}$  nicht invertierbar. Wir lösen das System der Normalgleichungen mit dem Eliminationsverfahren:

```
> R=rref(cbind(t(X)%*%X,t(X)%*%FIT))$echelon.form
> R
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
[1,] 1 0 0 0 0 0 0 0 0 0 1 128.53337
[2,] 0 1 0 0 0 0 0 0 0 0 1 151.92424
[3,] 0 0 1 0 0 0 0 0 0 0 1 166.95633
[4,] 0 0 0 1 0 0 0 0 0 0 1 175.47269
[5,] 0 0 0 0 1 0 0 0 0 0 1 207.33978
[6,] 0 0 0 0 0 0 1 0 0 0 -1 -46.01699
[7,] 0 0 0 0 0 0 0 1 0 0 -1 -43.14232
[8,] 0 0 0 0 0 0 0 0 1 0 -1 -31.59877
[9,] 0 0 0 0 0 0 0 0 0 1 -1 -26.98613
[10,] 0 0 0 0 0 0 0 0 0 0 0 0.00000
```

Da uns eine spezielle Lösung des Gleichungssystems genügt, setzen wir  $\nu_5 = 0$ , und erhalten

```
> mu.raw=R[1:5,11]
```

```
> mu.raw
[1] 128.5334 151.9242 166.9563 175.4727 207.3398
> nu.raw=R[6:10,11]
> nu.raw
[1] -46.01699 -43.14232 -31.59877 -26.98613 0.00000
```

Die von unserem Modell angepassten Zellmittelwerte lauten nun

```
> my.fitted=outer(mu.raw,nu.raw,"+")
> my.fitted
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 82.51638  85.39105  96.9346 101.5472 128.5334
[2,] 105.90725 108.78191 120.3255 124.9381 151.9242
[3,] 120.93934 123.81401 135.3576 139.9702 166.9563
[4,] 129.45570 132.33036 143.8739 148.4866 175.4727
[5,] 161.32279 164.19745 175.7410 180.3536 207.3398
```

Zu unserer Bestätigung sehen wir uns an, wie die angepassten Zellmittelwerte lauten, wenn wir die Funktion **lm** verwenden:

```
> tapply(fitted(lm(FIT~MTE+FTE)),list(MTE,FTE),mean)
      0      1      2      3      4
0 82.51638  85.39105  96.9346 101.5472 128.5334
1 105.90725 108.78191 120.3255 124.9381 151.9242
2 120.93934 123.81401 135.3576 139.9702 166.9563
3 129.45570 132.33036 143.8739 148.4866 175.4727
4          NA 164.19745 175.7410 180.3536 207.3398
```

Das sind die gleichen Werte. Lediglich in der leeren Zelle liefert uns R keinen Wert.

Um zu interpretierbaren Größen zu gelangen, ziehen wir den Durchschnitt aller Zellmittelwerte heraus:

```
> mm=mean(my.fitted)
> mu=mu.raw-mean(mu.raw)
> nu=nu.raw-mean(nu.raw)
> mm
[1] 136.4964
> mu
[1] -37.5119074 -14.1210450 0.9110499 9.4274058 41.2944967
> nu
[1] -16.468148 -13.593481 -2.049925 2.562712 29.548841
> mm+outer(mu,nu,"+")
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 82.51638  85.39105  96.9346 101.5472 128.5334
[2,] 105.90725 108.78191 120.3255 124.9381 151.9242
[3,] 120.93934 123.81401 135.3576 139.9702 166.9563
[4,] 129.45570 132.33036 143.8739 148.4866 175.4727
[5,] 161.32279 164.19745 175.7410 180.3536 207.3398
```

---

Die Beurteilung des Erklärungswertes des additiven Modells und die Modellselektion erfolgen dann wie gewohnt.

(5.62) BEISPIEL:

Wir passen die Modelle an:

```
> fm.0=lm(FIT~1)
> fm.1=lm(FIT~MTE)
> fm.2=lm(FIT~FTE)
> fm.12=lm(FIT~MTE+FTE)
```

Der Modellvergleich

```
> anova(fm.0, fm.1, fm.12)
Analysis of Variance Table
```

Model 1: FIT ~ 1

Model 2: FIT ~ MTE

Model 3: FIT ~ MTE + FTE

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	6031071				
2	1291	5026224	4	1004847	67.155	< 2.2e-16 ***
3	1287	4814370	4	211854	14.159	2.595e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

und

```
> anova(fm.0, fm.2, fm.12)
Analysis of Variance Table
```

Model 1: FIT ~ 1

Model 2: FIT ~ FTE

Model 3: FIT ~ MTE + FTE

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1295	6031071				
2	1291	5104021	4	927050	61.956	< 2.2e-16 ***
3	1287	4814370	4	289651	19.358	1.765e-15 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

zeigt uns, dass das additive Modell nicht weiter vereinfachbar ist.

---

## Modelle mit Wechselwirkung

---

Das Modell mit Wechselwirkung besteht einfach darin, dass für jede Zelle der Kreuztabelle der Zellmittelwert als Parameter verwendet wird:

$$y_i = m_{kl} + \epsilon_i \quad \text{wenn } i \in G_k \cap H_l$$

Die Modellanpassung ist daher sehr einfach und liefert als optimale Parameterschätzungen die empirischen Zellmittelwerte.

---

(5.63) BEISPIEL:

In unserem laufenden Beispiel erhalten wir

```
> fm.lw2=lm(FIT~MTE*FTE)
> anova(fm.l2, fm.lw2)
Analysis of Variance Table

Model 1: FIT ~ MTE + FTE
Model 2: FIT ~ MTE * FTE
  Res.Df    RSS    Df Sum of Sq      F Pr(>F)
1   1287 4814370
2   1272 4779787   15    34582 0.6135 0.866
```

Das Modell mit Wechselwirkung besitzt also keinen signifikanten Erklärungswert über das additive Modell hinaus.

---

(5.64) BEISPIEL:

Im folgenden Beispiel ist eine geringfügige Wechselwirkung zu bemerken:

```
> anova(lm(CTRA~MTE+CBSEX), lm(CTRA~MTE*CBSEX))
Analysis of Variance Table

Model 1: CTRA ~ MTE + CBSEX
Model 2: CTRA ~ MTE * CBSEX
  Res.Df    RSS    Df Sum of Sq      F Pr(>F)
1   1290 111673
2   1286 110818    4      854 2.4789 0.0424 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wir wollen uns ansehen, worin diese Wechselwirkung besteht. Zu diesem Zweck eliminieren wir aus der Kreuztabelle

```
> t=tapply(CTRA, list(CBSEX, MTE), mean)
> t
      0      1      2      3      4
0 23.0625 27.33333 27.98755 32.98020 37.84722
1 20.6000 24.94643 28.84980 30.80769 35.01361
```

alle Zeilen- und Spalteneffekte:

```
> mean.rows=apply(t, 1, mean)
> mean.cols=apply(t, 2, mean)
> mean.rows
      0      1
29.84216 28.04351
> mean.cols
      0      1      2      3      4
21.83125 26.13988 28.41868 31.89395 36.43041
> interaction=t-outer(mean.rows, mean.cols, "+")+mean(t)
```

```
> interaction
      0          1          2          3          4
0  0.3319223  0.2941247 -1.330453  0.1869252  0.5174807
1 -0.3319223 -0.2941247  1.330453 -0.1869252 -0.5174807
```

Die Wechselwirkung besteht offenbar darin, dass die Schwäche der Mädchen bei CTRA im Fall von MTE=2 nicht vorhanden ist.

## Orthogonale Versuchspläne

\*\*\*\*\* IN ARBEIT \*\*\*\*\*

### 5.4.4 Aufgaben

(5.65) **Aufgabe:** Untersuchen Sie im Datensatz **statlab** die Abhängigkeit der Variablen FIT von den Variablen MBAG und FBAG, ohne und mit Wechselwirkung. Interpretieren Sie gegebenenfalls eine signifikante Wechselwirkung in Worten.

(5.66) **Aufgabe:** Untersuchen Sie im Datensatz **statlab** die Abhängigkeit der Variablen CTHGHT von den Variablen MTHGHT und FTHGHT, ohne und mit Wechselwirkung. Interpretieren Sie gegebenenfalls eine signifikante Wechselwirkung in Worten.

(5.67) **Aufgabe:** Untersuchen Sie im Datensatz **statlab** die Abhängigkeit der Variablen CBWGT von den Variablen MTWGT und MTHGHT, ohne und mit Wechselwirkung. Interpretieren Sie gegebenenfalls eine signifikante Wechselwirkung in Worten.

(5.68) **Aufgabe:** Untersuchen Sie im Datensatz **kreditscoring** die Abhängigkeit der Laufzeit vom Alter des Kreditnehmers. Schalten Sie dabei den Einfluss der Kredithöhe aus.

(5.69) **Aufgabe:** Untersuchen Sie im Datensatz **kreditscoring** die Abhängigkeit der Laufzeit vom Beruf des Kreditnehmers. Schalten Sie dabei den Einfluss der Kredithöhe aus.

(5.70) **Aufgabe:** Wird im Datensatz **kreditscoring** der Zusammenhang zwischen Kredithöhe und Alter vom Beruf beeinflusst ? (Diagramm)

(5.71) **Aufgabe:** Wird im Datensatz **kreditscoring** der Zusammenhang zwischen Kredithöhe und Laufzeit vom Beruf beeinflusst ? (Diagramm)

(5.72) **Aufgabe:** Passen Sie für den Datensatz **kreditscoring** ein additives Modell an und berechnen Sie die Parameter:  $\text{ALTER} \sim \text{MORAL} + \text{BERUF}$ .

(5.73) **Aufgabe:** Passen Sie für den Datensatz **kreditscoring** ein additives Modell an und berechnen Sie die Parameter:  $\text{LAUFZ} \sim \text{MORAL} + \text{VERM}$ .

(5.74) **Aufgabe:** Führen Sie eine Modellselektion durch: Für den Datensatz **kreditscoring** und das saturierte Modell  $\text{ALTER} \sim \text{MORAL} + \text{BERUF}$ . Bestimmen Sie die wesentlichen Parameter des optimalen Modells.

(5.75) **Aufgabe:** Führen Sie eine Modellselektion durch: Für den Datensatz **kreditscoring** und das saturierte Modell  $\text{LAUFZ} \sim \text{MORAL} + \text{VERM}$ . Bestimmen Sie die wesentlichen Parameter des optimalen Modells.

(5.76) TESTFRAGEN

1. Erklären Sie den Unterschied zwischen additiven Modellen und Modellen mit Wechselwirkung.
2. Formulieren Sie für zwei quantitative Prädiktoren das additive lineare Regressionsmodell und das Modell mit Wechselwirkung.
3. Was versteht man unter einer Kovarianzanalyse ?
4. Formulieren Sie für zwei Prädiktoren von gemischtem Variablentyp das additive lineare Regressionsmodell und das Modell mit Wechselwirkung.
5. Auf welche Weise kann man die Abhängigkeit einer Regressionsgeraden von einem qualitativen Faktor modellieren ?
6. Wie interpretiert man den Parameter der Wechselwirkung zwischen einem dichotomen und einem quantitativen Prädiktor ?
7. Formulieren Sie für zwei qualitative Prädiktoren das additive lineare Regressionsmodell und das Modell mit Wechselwirkung.
8. Was versteht man unter zweifacher Varianzanalyse ?
9. Erklären Sie in Worten die Begriffe Haupteffekt und Wechselwirkungseffekt bei der zweifachen Varianzanalyse.

## 6

## Quadratische Funktionen

## 6.1 Funktionen von mehreren Variablen

## 6.1.1 Grundlegende Definitionen

Funktionen von mehreren Variablen sind Zuordnungsvorschriften der Form

$$f: \mathbb{R}^n \rightarrow \mathbb{R}: \mathbf{x} = (x_1, x_2, \dots, x_n) \mapsto f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$$

Der Definitionsbereich ist dabei eine Teilmenge von  $\mathbb{R}^n$ , der Wertebereich ist  $\mathbb{R}$ .

Wir behandeln in diesem Skriptum nur lineare und quadratische Funktionen. Der Grund dafür ist, dass die Analyse solche Funktionen mit Mitteln der linearen Algebra möglich ist.

**Lineare Funktionen**

(6.1) DEFINITION:

Unter einer **linearen Funktion** versteht man eine Funktion mit dem Funktionsterm

$$f(x_1, x_2, \dots, x_n) = b_1 x_1 + b_2 x_2 + \dots + b_n x_n + c$$

Ist  $c = 0$ , so spricht man von einer **homogenen** linearen Funktion. Eine homogene lineare Funktion hat die Eigenschaft

$$f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda f(x_1, x_2, \dots, x_n), \quad \lambda \in \mathbb{R}.$$

Man sagt, sie ist **homogen vom Grad 1**.

Bezeichnen wir mit  $\mathbf{b} = (b_1, b_2, \dots, b_n)^t$ , dann lässt sich die lineare Funktion in Matrixschreibweise angeben:

$$f(\mathbf{x}) = \mathbf{b} \cdot \mathbf{x} + c = \mathbf{b}^t \mathbf{x} + c$$

**Quadratische Funktionen**

Wir beginnen mit quadratischen Funktionen von zwei Variablen.

Eine **homogene quadratische Funktion** von zwei Variablen hat einen Funktionsterm der Form

$$f(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

Ein solcher Funktionsterm läßt sich immer schreiben als

$$f(x_1, x_2) = (x_1, x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^t \mathbf{A} \mathbf{x}$$

Die dabei auftretende Matrix  $\mathbf{A}$  ist grundsätzlich eine symmetrische Matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

(6.2) BEISPIEL:

Die quadratische Funktion

$$f(x_1, x_2) = 3x_1^2 - 8x_1x_2 + x_2^2$$

läßt sich schreiben als

$$f(x_1, x_2) = (x_1, x_2) \begin{pmatrix} 3 & -4 \\ -4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^t \mathbf{A} \mathbf{x}$$

wobei

$$\mathbf{A} = \begin{pmatrix} 3 & -4 \\ -4 & 1 \end{pmatrix}$$

Eine solche homogene quadratische Funktion hat die die Eigenschaft

$$f(\lambda \mathbf{x}) = (\lambda \mathbf{x})^t \mathbf{A} (\lambda \mathbf{x}) = \lambda^2 \mathbf{x}^t \mathbf{A} \mathbf{x} = \lambda^2 f(\mathbf{x})$$

Man sagt, sie ist **homogen vom Grad 2**.

Bei einer allgemeinen quadratischen Funktionen kommen noch lineare Glieder hinzu:

(6.3) DEFINITION:

Unter einer **quadratischen Funktion** versteht man eine Funktion mit dem Funktionsterm

$$f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{b}^t \mathbf{x} + c.$$

## 6.1.2 Graphische Darstellungen

Wir unterscheiden grundsätzlich zwischen drei Darstellungsformen für Funktionen von mehreren Variablen: **Funktionsgraph**, **Niveauliniendiagramm** und **Schnittliniendiagramm**.

### Funktionsgraphen

Ist  $y = f(\mathbf{x})$  ein Funktionsterm, dann nennt man die Punktmenge

$$G_f = \{(\mathbf{x}, f(\mathbf{x}))\}$$

den **Funktionsgraph** von  $f$ . Wenn  $\mathbf{x} \in \mathbb{R}^n$ , dann ist  $G_f \subseteq \mathbb{R}^{n+1}$ . Daher ist eine graphische Darstellung nur möglich, wenn  $n \leq 2$ .

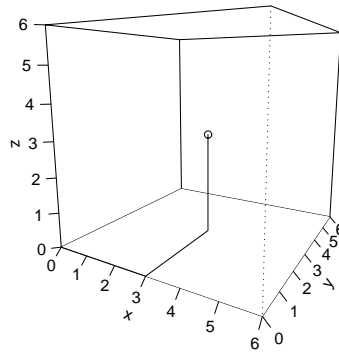


Abbildung 6.1: Darstellung des Punktes (3, 3, 3)

Wir werden also nur die Funktionsgraphen von Funktionen mit zwei Variablen graphisch darstellen können. Man verwendet dabei eine andere Bezeichnungsweise für die auftretenden Variablen:

$$z = f(x, y)$$

Ein Funktionsgraph kann immer nur an endlich vielen Punkten dargestellt werden. Man fertigt deshalb zunächst ein Gitter in der  $xy$ -Ebene an:

```
> x=seq(from=-3,to=3,by=0.5)
> y=x
```

Anschließend definiert man die Funktion, z.B. die lineare Funktion

$$f(x, y) = 2 * x - 3 * y + 1$$

```
ff=function(x,y) {2x-3y+1}
```

und berechnet den Funktionsgraph auf dem Gitter:

```
> z=outer(x,y,ff)
> z
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] ...
[1,]    4  2.5    1 -0.5   -2 -3.5   -5 -6.5   -8 ...
[2,]    5  3.5    2  0.5   -1 -2.5   -4 -5.5   -7 ...
[3,]    6  4.5    3  1.5    0 -1.5   -3 -4.5   -6 ...
...
```

Den Funktionsgraph zeichnet man dann durch

```
persp(x,y,z,ticktype="detailed",theta=20,phi=10)
```

Die beiden Optionen **theta** und **phi** geben die Winkel an, unter denen die perspektivische Darstellung erfolgt.

## Niveauliniendiagramm

---

Ist  $y = f(\mathbf{x})$  ein Funktionsterm, dann nennt man die Punktmenge

$$N_c = \{\mathbf{x} : f(\mathbf{x}) = c\}$$

die **Niveaumenge** von  $f$  zum Niveau  $c$ . Wenn  $\mathbf{x} \in \mathbb{R}^n$ , dann ist  $N_c \subseteq \mathbb{R}^n$ . Ixt  $n = 2$ , dann sind die Niveaumengen meist Niveaulinien.

Bei einem Niveauliniendiagramm zeichnet man die Niveaulinien für mehrere Niveaus gleichzeitig in das Diagramm ein. Dabei muss die Funktion  $z = f(x, y)$  zunächst wieder auf einem Gitter berechnet werden.

Mit unseren bisherigen Definitionen zeichnet

```
contour(x, y, z, nlevels=30)
```

ein Niveauliniendiagramm für 20 Niveaus. Es überrascht nicht, dass die Niveaulinien einer linearen Funktionen geradlinig verlaufen.

## Schnittliniendiagramm

---

Ein Niveauliniendiagramm ist bereits ein spezielles Schnittliniendiagramm: Die Fläche des Funktionsgraphen wird in der Höhe des Niveaus  $c$  durch einen horizontalen Schnitt aufgeschnitten und die entstehende Niveaulinie wird gezeichnet. Niveauliniendiagramme entstehen also durch bestimmte Schnitte:

- parallel zur  $xy$ -Ebene,
- orthogonal zur  $z$ -Achse (Blickrichtung der  $z$ -Achse).

Man kann natürlich in (genauer: orthogonal zu) jeder Blickrichtung Schnitte bilden. Wenn man das mit der  $x$ -Achse bzw. mit der  $y$ -Achse macht, entstehen die üblichen Schnittliniendiagramme.

(6.4) DEFINITION:

*Ist  $z = f(x, y)$  ein Funktionsterm, dann ist*

- 1. die Funktion  $x \mapsto f(x, c)$  die Schnittfunktion von  $f$  zum Schnitt (mit der Ebene)  $y = c$ ,*
- 2. die Funktion  $y \mapsto f(c, y)$  die Schnittfunktion von  $f$  zum Schnitt (mit der Ebene)  $x = c$ .*

Diese Schnittfunktionen sind Funktionen von einer Variablen und können daher ganz einfach gezeichnet werden.

In unserem laufenden Beispiel haben wir die Funktion  $z = f(x, y)$  auf einem Gitter mit den Gitterpunkten

```
> x
 [1] -3.0 -2.5 -2.0 -1.5 -1.0 -0.5  0.0  0.5  1.0  1.5
[11]  2.0  2.5  3.0
```

und

```
> y
 [1] -3.0 -2.5 -2.0 -1.5 -1.0 -0.5  0.0  0.5  1.0  1.5
[11]  2.0  2.5  3.0
```

gezeichnet. Die Funktionswerte haben wir in der Matrix

```
> z
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    4  2.5    1 -0.5   -2 -3.5   -5 -6.5   -8 -9.5 ...
[2,]    5  3.5    2  0.5   -1 -2.5   -4 -5.5   -7 -8.5 ...
[3,]    6  4.5    3  1.5    0 -1.5   -3 -4.5   -6 -7.5 ...
[4,]    7  5.5    4  2.5    1 -0.5   -2 -3.5   -5 -6.5 ...
[5,]    8  6.5    5  3.5    2  0.5   -1 -2.5   -4 -5.5 ...
[6,]    9  7.5    6  4.5    3  1.5    0 -1.5   -3 -4.5 ...
...

```

abgespeichert. Die Spalten der Matrix  $z$  sind also die  $y$ -Schnitte der Funktion  $f$  zu jene  $y$ -Werten, die auf dem  $y$ -Gitter liegen. Wir zeichnen diese Schnitte durch

```
matplot(x, z, type="l")
```

Analog dazu sind die Zeilen der Matrix  $z$  die  $x$ -Schnitte der Funktion  $f$  zu jenen  $x$ -Werten, die auf dem  $x$ -Gitter liegen. Wir zeichnen diese Schnitte durch

```
matplot(y, t(z), type="l")
```

### 6.1.3 Aufgaben

(6.5) **Aufgabe:** Stellen Sie die folgenden linearen Funktionen in Matrixschreibweise dar. Zeichnen Sie die Funktionsgraphen, die Niveauliniendiagramme und die Schnittliniendiagramme:

1.  $z = x + y$
2.  $z = 3x + 2y$
3.  $z = x - y$
4.  $z = 2x - 5y$
5.  $z = 2$
6.  $z = 2x$
7.  $z = 3y$

(6.6) **Aufgabe:** Stellen Sie die folgenden linearen Funktionen in Matrixschreibweise dar. Zeichnen Sie die Funktionsgraphen, die Niveauliniendiagramme und die Schnittliniendiagramme:

1.  $z = x^2 + y^2$  (Nach oben offenes Paraboloid, Niveaulinien sind Kreise.)
2.  $z = x^2 + 3y^2$  (Nach oben offenes Paraboloid, Niveaulinien sind achsenparallele Ellipsen.)
3.  $z = -x^2 - y^2$  (Nach unten offenes Paraboloid, Niveaulinien sind Kreise.)
4.  $z = -2x^2 - y^2$  (Nach unten offenes Paraboloid, Niveaulinien sind achsenparallele Ellipsen.)
5.  $z = x^2 - y^2$  (Sattelfläche, Niveaulinien sind achsenparallele Hyperbeln.)
6.  $z = x^2 - 2y^2$  (Sattelfläche, Niveaulinien sind achsenparallele Hyperbeln.)

(6.7) **Aufgabe:** Stellen Sie die folgenden linearen Funktionen in Matrixschreibweise dar. Zeichnen Sie die Funktionsgraphen, die Niveauliniendiagramme und die Schnittliniendiagramme. Interpretieren Sie den Typ des Funktionsgraphen und die Form der Niveaulinien:

1.  $z = xy$
2.  $z = -xy$
3.  $z = x^2$
4.  $z = -x^2$
5.  $z = y^2$
6.  $z = -y^2$
7.  $z = 2x^2 + 2xy + 3y^2$
8.  $z = x^2 + 4xy + 3y^2$
9.  $z = -x^2 + 2xy + y^2$
10.  $z = -x^2 + 2xy - 2y^2$

#### (6.8) TESTFRAGEN

1. Erklären Sie die Begriffe lineare und quadratische Funktion. Wie sehen die Funktionsterme in ausführlicher Schreibweise und in Matrixschreibweise aus ?
2. Was ist eine homogene Funktion ? Welche linearen Funktionen und welche quadratischen Funktionen sind homogen ?
3. Erklären Sie die wichtigsten graphischen Darstellungsmöglichkeiten für Funktionen mit mehreren Variablen.

## 6.2 Die Hauptachsentransformation

### 6.2.1 Orthogonale Matrizen

(6.9) DEFINITION:

Es seien  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$  Vektoren in  $\mathbb{R}^n$ . Diese Vektoren bilden ein **Orthonormalsystem (ONS)**, wenn

1. alle Vektoren paarweise zueinander orthogonal sind,
2. alle Vektoren normiert sind (die Länge 1 haben).

(6.10) BEISPIEL:

Die Vektoren

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

bilden ein ONS. Ebenso bilden die Vektoren

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{a}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

ein ONS. Die Vektoren

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

bilden kein ONS: Sie sind zwar orthogonal zueinander, aber nicht normiert.

(6.11) DEFINITION:

*Eine quadratische Matrix heisst **orthogonal**, wenn ihre Spalten ein ONS bilden.*

Bevor wir uns Beispielen zuwenden, sehen wir uns den Begriff der orthogonalen Matrix noch etwas genauer an.

(6.12) SATZ:

*Es sei  $T$  eine quadratische Matrix. Dann sind die folgenden Aussagen äquivalent:*

1.  $T$  ist eine orthogonale Matrix, dh. die Spalten von  $T$  bilden ein ONS.
2. Die Zeilen von  $T$  bilden ein ONS.
3. Die Transponierte  $T^t$  ist identisch mit der Inversen  $T^{-1}$ , dh.  $T^t T = E$ .
4. Die durch  $T$  definierte Transformation ist längentreu, dh.  $\|T\mathbf{x}\| = \|\mathbf{x}\|$  für alle  $\mathbf{x} \in \mathbb{R}^n$ .

BEGRÜNDUNG:

Wir sehen uns den Beweis dieses Satzes deshalb an, weil er mit unseren Mitteln leicht durchzuführen ist, und weil man daraus viel lernen kann.

(1)  $\implies$  (3): Wenn die Spalten von  $T$  ein ONS bilden, dann gilt

$$\mathbf{a}_i^t \mathbf{a}_j = \mathbf{a}_i \cdot \mathbf{a}_j = \begin{cases} 1 & \text{wenn } i = j, \\ 0 & \text{wenn } i \neq j \end{cases}$$

Bei der Matrixmultiplikation von  $T^t$  mit  $T$  werden gerade diese Multiplikationen ausgeführt. Daher entsteht dabei die Einheitmatrix, also  $T^t T = E$ .

(3)  $\implies$  (2) und (2)  $\implies$  (3): Wir wissen, dass die Inverse  $T^{-1}$  die Eigenschaften

$$T^{-1} T = T T^{-1} = E$$

besitzt. Daraus folgt nun, dass

$$T^t T = T T^t = E$$

Die zweite Gleichung besagt aber nichts anderes, als dass die Zeilen von  $T$  ein ONS bilden.

(1)  $\implies$  (4): Dies folgt ganz rasch aus

$$\|T\mathbf{x}\|^2 = \mathbf{x}^t T^t T \mathbf{x} = \mathbf{x}^t \mathbf{x} = \|\mathbf{x}\|^2$$

(4)  $\implies$  (2): Das ist etwas komplizierter. Aber es ist dafür der interessanteste Teil der ganzen Aussage: Er besagt, dass die einzigen Matrizen, die längentreue Transformationen bewirken, die orthogonalen Matrizen sind.

Aus

$$\|T(\mathbf{x} - \mathbf{y})\|^2 = \|\mathbf{x} - \mathbf{y}\|^2$$

folgt

$$\mathbf{x}^t T^t T \mathbf{x} - \mathbf{x}^t T^t T \mathbf{y} - \mathbf{y}^t T^t T \mathbf{x} + \mathbf{y}^t T^t T \mathbf{y} = \mathbf{x}^t \mathbf{x} - \mathbf{x}^t \mathbf{y} - \mathbf{y}^t \mathbf{x} + \mathbf{y}^t \mathbf{y}$$

Da wir die Eigenschaft (4) vorausgesetzt haben, heben sich jene Terme auf der linken und rechten Seite weg, die entweder nur  $x$  oder nur  $y$  enthalten. Es verbleibt die Gleichung

$$-x^t T^t T y - y^t T^t T x = -x^t y - y^t x$$

Da auf jeder der beiden Seiten ein Term gerade der transponierte des anderen ist, und weil es gleichzeitig  $1 \times 1$ -Matrizen sind, erhalten wir schließlich

$$x^t T^t T y = x^t y = x^t E y \quad \text{für alle } x, y$$

Wenn wir  $x = e_i$  und  $y = e_j$  setzen, ergibt sich auf jeder Seite das Element der Matrix mit dem Index  $ij$ . Daher sind die beiden Matrizen gleich.  $\square$

Die Aussagen (1)–(3) dienen nur dazu, die Eigenschaft einer Matrix, orthogonal zu sein, leichter überprüfen zu können. Die Aussage (4) ist dagegen von inhaltlicher Bedeutung.

## Orthogonale $2 \times 2$ -Matrizen

Im Fall  $n = 2$  sind die einzigen orthogonalen Matrizen diejenigen, die Drehungen oder Drehspiegelungen beschreiben. Das kann man theoretisch begründen oder experimentell veranschaulichen.

Wir beginnen mit der experimentellen Veranschaulichung. Zu diesem Zweck zeichnen wir ein Dreieck:

```
coord(xlim=c(-5,5),ylim=c(-5,5))
x1=c(-3,-3)
x2=c(2,-1)
x3=c(1,3)
lines(t(cbind(x1,x2,x3,x1)))
```

Auf die Eckpunkte des Dreiecks wollen wir eine orthogonale Matrix anwenden. Wir werden sehen, dass dadurch eine Drehung oder eine Drehspiegelung bewirkt wird.

Ein beliebige orthogonale Matrix

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

muss die Gleichungen

$$a^2 + c^2 = 1, \quad b^2 + d^2 = 1, \quad ab + cd = 0$$

erfüllen. Aus der dritten Gleichung folgt

$$\frac{a}{c} = -\frac{d}{b} := \lambda \quad \implies \quad a = \lambda c, \quad d = -\lambda b.$$

Dies setzen wir in die erste und zweite Gleichung ein und erhalten  $b^2 = c^2$ . Daher ist  $b = \pm c$ .

**Fall 1:**  $b = -c$ . Es gilt:  $c = \sqrt{1-a^2}$ ,  $b = -c$ ,  $d = \sqrt{1-b^2} = a$ . Dabei ist  $a$  irgendeine Zahl zwischen 0 und 1:

$$T = \begin{pmatrix} a & -\sqrt{1-a^2} \\ \sqrt{1-a^2} & a \end{pmatrix}$$

Es sei also z.B.

```
a=0.1
```

```
T=cbind(c(a,sqrt(1-a^2)), -c(sqrt(1-a^2), a))
```

Wir zeichnen nun das transformierte Dreieck:

```
lines(t(T%*%cbind(x1,x2,x3,x1)))
```

Offensichtlich wurde das Dreieck gedreht. Die Abstände zwischen den Eckpunkten haben sich nicht verändert. Dies liegt an der Längentreue der orthogonalen Transformation.

**Fall 2:**  $b = c$

$$\mathbf{T} = \begin{pmatrix} a & \sqrt{1-a^2} \\ \sqrt{1-a^2} & a \end{pmatrix}$$

Wir zeichnen wieder das transformierte Dreieck:

```
a=0.1
```

```
T=cbind(c(a,sqrt(1-a^2)), c(sqrt(1-a^2), a))
```

```
lines(t(T%*%cbind(x1,x2,x3,x1)))
```

Offensichtlich wurde das Dreieck nun gespiegelt. Die Abstände zwischen den Eckpunkten haben sich aber wieder nicht verändert. Dies liegt an der Längentreue der orthogonalen Transformation.

Man kann diese Experimente mit verschiedenen Werten für  $a$  durchführen. Dadurch ändern sich die Winkel der Drehungen bzw. Spiegelungen.

Um diese Erscheinungen theoretisch zu verstehen, muss man nur  $a = \cos \phi$  setzen. Dann ist  $\phi$  der Drehungswinkel bzw. der Winkel des Orthogonalvektors der Spiegelungsachse.

## 6.2.2 Diagonalisierung von symmetrischen Matrizen

Der folgende Satz enthält eine grundlegende Aussage über symmetrische Matrizen.

(6.13) SATZ:

*Es sei  $A$  eine symmetrische Matrix. Dann gibt es eine orthogonale Matrix  $T$ , sodass*

$$T^{-1}AT = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

*Die Zahlen  $\lambda_1, \lambda_2, \dots, \lambda_n$  heissen **Eigenwerte** von  $A$ . Sie sind bis auf die Reihenfolge eindeutig bestimmt. Die Spalten der orthogonalen Matrix  $T$  heissen **Eigenvektoren** von  $A$ .*

Sowohl der Beweis dieses Satzes als auch die numerische Berechnung der Eigenwerte und Eigenvektoren führt weit über das hinaus, was wir in diesem Skriptum behandeln können. Trotzdem können wir damit arbeiten.

### Numerische Diagonalisierung

Wir können mit R Eigenwerte und Eigenvektoren berechnen.

Es sei

```
A=cbind(c(2,1),c(1,3))
```

eine symmetrische Matrix. Die Funktion **eigen** liefert uns die Eigenwerte

```
> D=diag(eigen(A)$values)
> D
      [,1]      [,2]
[1,] 3.618034 0.000000
[2,] 0.000000 1.381966
```

und die Eigenvektoren

```
> T=eigen(A)$vectors
> T
      [,1]      [,2]
[1,] 0.5257311 0.8506508
[2,] 0.8506508 -0.5257311
```

Wir überprüfen nun die im Lehrsatz geforderten Eigenschaften der Diagonalisierung. Die Orthogonalität der Matrix  $T$  bestätigen wir durch

```
> round(t(T)%*%T,8)
      [,1] [,2]
[1,] 1 0
[2,] 0 1
> round(T%*%t(T),8)
      [,1] [,2]
[1,] 1 0
[2,] 0 1
```

Die Diagonalisierung erfolgt durch

```
> round(t(T)%*%A%*%T,8)
      [,1]      [,2]
[1,] 3.618034 0.000000
[2,] 0.000000 1.381966
```

Man kann auch wesentlich größere symmetrische Matrizen diagonalisieren. Wir erzeugen eine zufällige symmetrische Matrix durch

```
> R=matrix(rnorm(100),10,10)
> A=R+t(R)
> A
      [,1]      [,2]      [,3]      [,4]
[1,] -1.5434403 -1.4383492 -0.40692333 -0.42604214 ...
[2,] -1.4383492 -0.1176705 0.46552606 -1.27186802 ...
[3,] -0.4069233 0.4655261 -0.08122003 -2.17250684 ...
[4,] -0.4260421 -1.2718680 -2.17250684 1.97612846 ...
[5,] 1.1538595 0.3760915 -0.70620813 -0.01362711 ...
[6,] 1.4742792 -0.6847583 -0.73371974 1.63324947 ...
...
```

Die Eigenwerte sind

```
> eigen(A)$values
[1] 5.4041198 4.9686435 4.3446504 1.8660153 0.6750763
```

[6] -0.4785932 -1.4251390 -2.1917073 -3.3965868 -4.9973265

Der von R verwendete Algorithmus gibt die Eigenwerte stets der Größe nach geordnet an. Wir sehen, dass die Eigenwerte sowohl positiv als auch negativ sein können. Das Vorzeichen der Eigenwerte wird im folgenden eine wichtige Rolle spielen.

### 6.2.3 Anwendung auf quadratische Funktionen

(6.14) DEFINITION:

Eine homogene quadratische Funktion befindet sich in **Normalform**, wenn ihr Funktionsterm nur aus quadratischen Gliedern besteht, dh. wenn

$$f(x_1, x_2, \dots, x_n) = a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{nn}x_n^2$$

Wenn sich eine homogene quadratische Funktion in Normalform befindet, dann kann man wichtige geometrische Eigenschaften des Funktionsgraphen bereits am Funktionsterm ablesen (siehe Abschnitt ...)

Wir werden nun sehen, dass man eine homogene quadratische Funktion, die sich nicht in Normalform befindet, durch eine lineare Transformation auf Normalform bringen kann. Das technische Hilfsmittel dazu ist die Diagonalisierung von symmetrischen Matrizen.

#### Einführung

Wir wollen an einem Beispiel den Grundgedanken der Hauptachsentransformation für quadratische Funktionen erläutern.

Wir untersuchen die quadratische Funktion

$$f(x_1, x_2) = x_1^2 + 4x_1x_2 + 3x_2^2$$

Wir nehmen nun das Ergebnis der nachfolgenden Berechnungen vorweg, damit klar ist, worauf wir hinauswollen: Wir werden zeigen, dass für diese quadratische Funktion die folgende Darstellung möglich ist:

$$\begin{aligned} f(x_1, x_2) &= x_1^2 + 4x_1x_2 + 3x_2^2 \\ &= 4.326(0.526x_1 + 0.851x_2)^2 - 0.236(0.851x_1 - 0.526x_2)^2 \end{aligned}$$

Diese Umformung heisst **Hauptachsentransformation**.

Zwei Fragen drängen sich auf:

1. **Wozu soll diese Umformung gut sein ?**
2. **Wie kommt man zu der Darstellung ?**

Zunächst beantworten wir die erste Frage. Wir vereinfachen die Darstellung, indem wir für die Ausdrücke in den Klammern Abkürzungen einsetzen („Substitution“):

$$f(x_1, x_2) = 4.326 \underbrace{(0.526x_1 + 0.851x_2)}_{y_1}^2 - 0.236 \underbrace{(0.851x_1 - 0.526x_2)}_{y_2}^2$$

Dadurch erhalten wir

$$f(x_1, x_2) = 4.326y_1^2 - 0.236y_2^2$$

Man kann die Substitution

$$\begin{aligned}y_1 &= 0.526x_1 + 0.851x_2 \\y_2 &= 0.851x_1 - 0.526x_2\end{aligned}$$

als „Koordinatentransformation“ interpretieren. Diese „Koordinatentransformation“ hat bewirkt, dass die quadratische Funktion eine Normalform angenommen hat. Wir können nun voraussagen, wie der Funktionsgraph dieser Normalform aussehen wird: Es handelt sich um eine Sattelfläche, weil die Vorzeichen der Koeffizienten verschieden sind.

In der geometrischen Darstellung bewirkt die „Koordinatentransformation“ eine Drehung oder eine Spiegelung. Davon können wir uns leicht überzeugen, indem wir die Niveauliniendiagramme zeichnen:

```
> x=seq(from=-3,to=3,by=0.5)
> y=x
> z=outer(x,y,function(x,y) x^2+4*x*y+3*y^2)
> contour(x,y,z)
> z=outer(x,y,function(x,y) 4.236*x^2-0.236*y^2)
> contour(x,y,z)
```

Also liefert uns die Deutung der Normalform zugleich die Information darüber, wie der Funktionsgraph der ursprünglichen quadratischen Funktion aussieht. Das ist die Antwort auf die erste der beiden Fragen:

**Die Hauptachsentransformation hat den Zweck, eine homogene quadratische Funktion in eine Normalform zu bringen, damit man geometrische Eigenschaften des Funktionsgraphen daraus ablesen kann.**

Nun wenden wir uns der Beantwortung der zweiten Frage zu. Wir werden sehen, dass die rechnerische Durchführung der Hauptachsentransformation auf der Diagonalisierung von symmetrischen Matrizen beruht.

Unsere quadratische Funktion hat die Matrixdarstellung

$$f(x_1, x_2) = x_1^2 + 4x_1x_2 + 3x_2^2 = (x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mathbf{x}^t \mathbf{A} \mathbf{x}$$

Wir diagonalisieren nun die Matrix  $A$ :

```
> A=cbind(c(1,2),c(2,3))
> D=diag(eigen(A)$values)
> D
      [,1]      [,2]
[1,] 4.236068  0.000000
[2,] 0.000000 -0.236068
> T=eigen(A)$vectors
> T
      [,1]      [,2]
[1,] 0.5257311  0.8506508
[2,] 0.8506508 -0.5257311
```

Wir wissen nun, dass

$$\mathbf{T}^t \mathbf{A} \mathbf{T} = \mathbf{D} \quad \text{und} \quad \mathbf{A} = \mathbf{T} \mathbf{D} \mathbf{T}^t$$

Daher können wir unsere quadratische Funktion umformen:

$$f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} = \mathbf{x}^t (\mathbf{T} \mathbf{D} \mathbf{T}^t) \mathbf{x} = (\mathbf{T}^t \mathbf{x})^t \mathbf{D} (\mathbf{T}^t \mathbf{x})$$

Indem wir  $\mathbf{y} = \mathbf{T}^t \mathbf{x}$  substituieren, erhalten wir

$$f(\mathbf{x}) = \mathbf{y}^t \mathbf{D} \mathbf{y} = 4.326y_1^2 - 0.236y_2^2$$

Das ist genau das Ergebnis der Hauptachsentransformation. Und damit ist die zweite Frage beantwortet.

## Der allgemeine Fall

---

Wir fassen nun den Vorgang der Hauptachsentransformation allgemein zusammen:

Gegeben sei eine homogene quadratische Funktion  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$  mit der symmetrischen Matrix  $\mathbf{A}$ .

**1. Schritt:** Die Matrix  $\mathbf{A}$  wird diagonalisiert.

Wir bezeichnen die Eigenwerte mit  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Die Eigenvektoren werden als Spalten der Matrix  $\mathbf{T}$  zusammengefasst. Die Diagonalmatrix der Eigenwerte bezeichnen wir mit  $\mathbf{D}$ .

**2. Schritt:** Wir substituieren in der quadratischen Funktion  $\mathbf{x} = \mathbf{T} \mathbf{y}$ , dh. wir definieren  $\mathbf{y} := \mathbf{T}^t \mathbf{x}$ .

**3. Schritt:** Als Ergebnis erhalten wir die Normalform

$$f(x_1, x_2, \dots, x_n) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$$

Die Koordinatentransformation  $\mathbf{x} = \mathbf{T} \mathbf{y}$  ist eine orthogonale Transformation, dh. sie bewirkt geometrisch eine Drehung oder Drehspiegelung des Koordinatensystems. Daher sind für die ursprüngliche quadratische Form und ihre Normalform jene Eigenschaften gleich, die sich unter Drehungen oder Drehspiegelungen nicht ändern.

Im nächsten Abschnitt lernen wir zwei wichtige Eigenschaften dieser Art kennen.

## 6.2.4 Aufgaben

### (6.15) TESTFRAGEN

1. Was versteht man unter einem Orthonormalsystem ? Geben Sie Beispiele für Orthonormalsysteme an.
2. Was ist eine orthogonale Matrix ? Wie überprüft man die Orthogonalität einer Matrix ? Worin besteht die geometrische Interpretation einer orthogonalen Matrix ?
3. Auf welche Weise kann man alle orthogonalen 2x2-Matrizen erzeugen ?
4. Wie lautet der Hauptsatz über die Diagonalisierung von symmetrischen Matrizen ?
5. Was versteht man unter Eigenwerten und Eigenvektoren ?
6. Wann befindet sich eine quadratische Funktion in Normalform ?

7. Was ist die Hauptachsentransformation einer quadratischen Funktion ? Wozu dient sie ?
8. Wie führt man die Hauptachsentransformation einer quadratischen Funktion durch ?
9. Was bewirkt die Hauptachsentransformation einer quadratischen Funktion für das geometrische Bild des Funktionsgraphen ?

## 6.3 Konvexe und konkave Funktionen

### 6.3.1 Grundlegende Definitionen

(6.16) DEFINITION:

Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  heisst **konvex**, wenn für je zwei Punkte  $\mathbf{x}_1$  und  $\mathbf{x}_2$  die Ungleichung

$$f\left(\frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)\right) \leq \frac{1}{2}(f(\mathbf{x}_1) + f(\mathbf{x}_2))$$

erfüllt ist.

Anschaulich bedeutet Konvexität, dass der Funktionsgraph stets unterhalb der Verbindungsstrecke zweier Funktionswerte liegt.

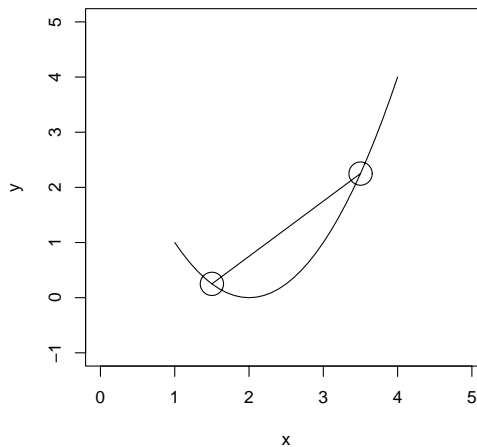


Abbildung 6.2: Konvexe Funktion

(6.17) DEFINITION:

Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  heisst **konkav**, wenn für je zwei Punkte  $\mathbf{x}_1$  und  $\mathbf{x}_2$  die Ungleichung

$$f\left(\frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)\right) \geq \frac{1}{2}(f(\mathbf{x}_1) + f(\mathbf{x}_2))$$

erfüllt ist.

Anschaulich bedeutet Konkavität, dass der Funktionsgraph stets oberhalb der Verbindungsstrecke zweier Funktionswerte liegt.

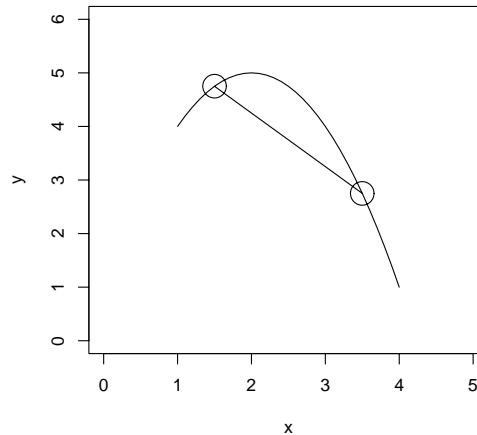


Abbildung 6.3: Konkave Funktion

Für die Behandlung von Optimierungsaufgaben ist es wichtig zu wissen, ob eine Funktion konvex oder konkav oder keines von beiden ist. Wir verschaffen uns daher einen einfachen Überblick über typische konvexe und konkave Funktionen.

- **Eine lineare Funktion ist immer sowohl konvex als auch konkav.**

Das ist völlig klar, denn für lineare Funktionen gelten sogar die Gleichungen

$$f\left(\frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)\right) = \frac{1}{2}(f(\mathbf{x}_1) + f(\mathbf{x}_2))$$

- **Eine quadratische Funktion  $f(x) = ax^2$  ist konvex, wenn  $a > 0$  und konkav, wenn  $a < 0$ .**

Das ist anschaulich klar, denn die Funktionsgraphen von solchen quadratischen Funktionen können wir zeichnen.

- **Wenn man konvexe Funktionen addiert, entsteht wieder eine konvexe Funktion.**

**Wenn man konkave Funktionen addiert, entsteht wieder eine konkave Funktion.**

Diese Aussagen folgen aus dem Umstand, dass die Ungleichungen, die Konvexität bzw. Konkavität definieren, bei Addition erhalten bleiben.

## 6.3.2 Quadratische Funktionen

Mit diesen einfachen und plausiblen Tatsachen sind wir in der Lage, alle quadratischen Funktionen dahingehend zu beurteilen, ob sie konvex oder konkav sind.

Wir beginnen mit homogenen quadratischen Funktionen in Normalform.

(6.18) SATZ:

Es sei

$$f(x_1, x_2, \dots, x_n) = a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{nn}x_n^2$$

eine homogene quadratische Funktion in Normalform. Dann gilt:

1.  $f$  ist konvex, wenn  $a_{ii} \geq 0$  für  $i = 1, 2, \dots, n$ .
2.  $f$  ist konkav, wenn  $a_{ii} \leq 0$  für  $i = 1, 2, \dots, n$ .

Dieser Satz folgt unmittelbar aus unseren einfachen Regeln. Man kann ihn dahingehend ergänzen, dass die Funktion  $f$  weder konvex noch konkav ist, wenn zwei von Null verschiedene Koeffizienten unterschiedliches Vorzeichen besitzen.

Wir sind nun in der Lage, zu beurteilen, ob eine homogene quadratische Funktion konvex oder konkav ist, wenn sie sich in Normalform befindet, dh. wenn ihr Funktionsterm keine gemischten Glieder besitzt.

Diese Möglichkeit ist aber nicht nur auf Normalformen beschränkt. Uns steht ja die Hauptachsentransformation zu Verfügung. Die Hauptachsentransformation besagt, dass wir jede homogene quadratische Funktion in der Form

$$f(x_1, x_2, \dots, x_n) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$$

schreiben können. Dabei sind die Variablen  $y_i$  lineare Funktionen der ursprünglichen Variablen  $x_i$ , die eine Drehung oder Drehspiegelung des Koordinatensystems bewirken. Bei einer Drehung oder Drehspiegelung bleiben Eigenschaften wie Konvexität oder Konkavität des Funktionsgraphen unverändert. Daher können wir aus der Normalform, die durch die Hauptachsentransformation entsteht, ablesen, ob eine quadratische Funktion konvex oder konkav oder keines von beiden ist.

(6.19) SATZ:

Ob ein homogene quadratische Funktion  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$  konvex oder konkav ist, hängt von den Eigenwerten der Matrix  $\mathbf{A}$  ab:

1.  $f$  ist genau dann konvex, wenn alle Eigenwerte nichtnegativ sind.
2.  $f$  ist genau dann konkav, wenn alle Eigenwerte nichtnegativ sind.

Wie steht es aber nun mit inhomogenen quadratischen Funktionen? Die Antwort auf diese Frage ist sehr einfach.

(6.20) SATZ:

Eine inhomogene quadratische Funktion  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{b}^t \mathbf{x} + c$  ist genau dann konvex bzw. konkav, wenn die zugehörige homogene quadratische Funktion  $\mathbf{x}^t \mathbf{A} \mathbf{x}$  konvex bzw. konkav ist.

Die Begründung für dafür ist sehr einfach: Die inhomogene Funktion unterscheidet sich von der homogenen Funktion nur durch den linearen Funktionsterm  $\mathbf{b}^t \mathbf{x} + c$ . Diese lineare Funktion ist sowohl konvex als auch konkav. Ihre Addition oder Subtraktion ändert also nichts an der Konvexität bzw. Konkavität der einen wie der anderen quadratischen Funktion.

### 6.3.3 Aufgaben

(6.21) **Aufgabe:** Für die folgende quadratischen Funktionen beantworte man die Fragen:

1. Wie lautet die Matrixdarstellung ?
2. Man führe die Hauptachsentransformation durch.
3. Wie lauten die Eigenwerte und die Eigenvektoren ?
4. Wie lautet die Normalform ?
5. Mit welcher orthogonalen Transformation erhält man die Normalform ?
6. Ist die quadratische Funktion konvex oder konkav ?
7. Zeichnen Sie den Funktionsgraph ? Welche Form hat der Funktionsgraph ?
8. Zeichnen Sie die Niveauliniendiagramme der quadratischen Funktion und ihrer Normalform.

Angaben:

1.  $f(x_1, x_2) = -4x_1^2 - 4x_1x_2 - x_2^2$
2.  $f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + 2x_2^2$
3.  $f(x_1, x_2) = 3x_1^2 - 4x_1x_2 + 2x_2^2$
4.  $f(x_1, x_2) = 2x_1^2 + 8x_1x_2 + 4x_2^2$
5.  $f(x_1, x_2) = -4x_1^2 - 6x_1x_2 - 2x_2^2$
6.  $f(x_1, x_2) = -3x_1^2 - 4x_1x_2$
7.  $f(x_1, x_2) = 3x_1^2 + 8x_1x_2 + 3x_2^2$
8.  $f(x_1, x_2) = -3x_1^2 + 10x_1x_2 + 1x_2^2$
9.  $f(x_1, x_2) = -4x_1^2 - 4x_1x_2 - 2x_2^2$
10.  $f(x_1, x_2) = -2x_1^2 + 4x_1x_2 - 4x_2^2$
11.  $f(x_1, x_2) = 3x_1^2 + 2x_1x_2 + 4x_2^2$
12.  $f(x_1, x_2) = 5x_1^2 + 6x_1x_2 + 4x_2^2$
13.  $f(x_1, x_2) = 5x_1^2 - 6x_1x_2 + 4x_2^2$
14.  $f(x_1, x_2) = -2x_1^2 + 2x_1x_2 - 2x_2^2$
15.  $f(x_1, x_2) = -2x_1^2 + 2x_1x_2 - 4x_2^2$
16.  $f(x_1, x_2) = 3x_1^2 + 6x_1x_2 + 3x_2^2$
17.  $f(x_1, x_2) = 4x_1^2 + 4x_1x_2 + x_2^2$
18.  $f(x_1, x_2) = -x_1^2 + 2x_1x_2 - x_2^2$
19.  $f(x_1, x_2) = -4x_1^2 + 2x_1x_2 - 3x_2^2$
20.  $f(x_1, x_2) = -4x_1^2 + 4x_1x_2 - x_2^2$

#### (6.22) TESTFRAGEN

1. Was versteht man unter einer konvexen Funktion ? Worin besteht die anschauliche Deutung dieses Begriffs ?
2. Was versteht man unter einer konkaven Funktion ? Worin besteht die anschauliche Deutung dieses Begriffs ?

3. Unter welchen Umständen sind lineare Funktionen konvex bzw. konkav ?
4. Wie erkennt man an einer quadratischen Funktion in Normalform, ob sie konvex oder konkav ist ?
5. Wie erkennt man an einer beliebigen quadratischen Funktion, ob sie konvex oder konkav ist ?

## 7

## Quadratische Optimierung

## 7.1 Grundlagen aus der Analysis

## 7.1.1 Einleitung

Wenn wir von einer differenzierbaren Funktion  $y = f(x)$ , die nur eine Variable besitzt, eine Maximalstelle oder eine Minimalstelle finden wollen, gehen wir in der Regel so vor:

1. Wir bilden die **Ableitung**  $f'(x)$ .
2. Wir suchen die **kritischen Punkte**, das sind die Punkte  $x$ , für die die Ableitung den Wert  $f'(x) = 0$  hat.
3. Wir beurteilen die kritischen Punkte dahingehend, ob die Funktion  $f$  dort tatsächlich ein Minimum oder ein Maximum annimmt.

Der Hintergrund dieser Vorgangsweise besteht darin, dass bei einer differenzierbaren Funktion jede Minimalstelle und jede Maximalstelle zugleich ein kritischer Punkt sein muss. Die anschauliche Begründung dafür ist unmittelbar einsichtig: An einer Minimalstelle oder einer Maximalstelle (einer differenzierbaren Funktion im Innern ihres Definitionsbereichs) muss die Tangente an den Funktionsgraphen waagrecht verlaufen. Da die Ableitung der Funktion zugleich die Steigung der Tangente ist, muss daher die Ableitung gleich Null sein.

Will man von einer Funktion  $y = f(x_1, x_2, \dots, x_n)$  mit mehreren Variablen das Minimum oder Maximum finden, dann geht man ganz ähnlich vor: Man sucht zunächst kritische Punkte, indem man etwas, was dem Begriff der Ableitung entspricht, gleich Null setzt. Anschließend beurteilt man die gewonnenen kritischen Punkte dahingehend, ob die Funktion dort ein Minimum oder Maximum besitzt.

Wir müssen daher zuerst klären, was bei einer Funktion mit mehreren Variablen dem Begriff der Ableitung entspricht. Dieses Thema erschöpfend zu behandeln, geht aber über den Rahmen dieses Textes hinaus. Wir beschränken uns daher auf das, was für das Verständnis unseres Themas, nämlich der Optimierung von quadratischen Funktionen, unbedingt nötig ist.

## 7.1.2 Richtungsableitung

Der grundlegende Begriff für alles weitere ist der Begriff der **Richtungsableitung**. Anschaulich gesprochen, geht es dabei um folgendes.

Es sei  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , eine Funktion mit mehreren Variablen.

1. Man legt einen Punkt  $\mathbf{x}_0$  und einen Richtungsvektor  $\mathbf{v}$  fest. Diese beiden Größen definieren eine Gerade

$$t \mapsto \mathbf{x}_0 + t\mathbf{v}, \quad t \in \mathbb{R}$$

im Definitionsbereich der Funktion  $f$ .

2. Anschließend untersucht man die Funktion  $f$  entlang dieser Geraden:

$$t \mapsto f(\mathbf{x}_0 + t\mathbf{v}) =: g(t)$$

Die entstehende Funktion  $g(t)$  ist nun eine Funktion mit einer einzigen Variablen  $t$ , und sie gibt an, welche Werte unsere ursprüngliche Funktion  $f$  entlang der Geraden  $t \mapsto \mathbf{x}_0 + t\mathbf{v}$  annimmt.

3. Unter der Richtungsableitung von  $f$  an der Stelle  $\mathbf{x}_0$  in Richtung  $\mathbf{v}$  versteht man nun einfach die Ableitung der Funktion  $g$  an der Stelle  $t = 0$ .

Diese Richtungsableitung gibt die Veränderungsrate der Funktion  $f$  an, die entsteht, wenn man sich vom Punkt  $\mathbf{x}_0$  in Richtung  $\mathbf{v}$  weiterbewegt.

(7.1) DEFINITION:

*Es sei  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , eine Funktion. Unter der Richtungsableitung von  $f$  an der Stelle  $\mathbf{x}_0$  in Richtung  $\mathbf{v}$  versteht man die Ableitung der Funktion*

$$g(t) = f(\mathbf{x}_0 + t\mathbf{v}), \quad t \in \mathbb{R},$$

*an der Stelle  $t = 0$ , also*

$$\left. \frac{d}{dt} g(t) \right|_{t=0} = \left. \frac{d}{dt} f(\mathbf{x}_0 + t\mathbf{v}) \right|_{t=0}$$

(7.2) BEISPIEL:

Wir betrachten die Funktion  $y = f(x_1, x_2) = 3x_1^2 - x_2^2$ . Wir wollen die Richtungsableitung an der Stelle  $\mathbf{x}_0 = (1, 2)^t$  in Richtung  $\mathbf{v} = (-1, 1)^t$  berechnen.

Die Gerade, entlang der wir die Funktion  $f$  untersuchen müssen, lautet

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 - t \\ 2 + t \end{pmatrix},$$

also

$$\begin{aligned} x_1 &= 1 - t \\ x_2 &= 2 + t \end{aligned}$$

Wir setzen diese Gerade nun in die Funktion  $f$  ein und erhalten so

$$g(t) = f(1 - t, 2 + t) = 3(1 - t)^2 - (2 + t)^2 = 2t^2 - 10t - 1$$

Daraus folgt  $g'(0) = -10$ . Also beträgt die gesuchte Richtungsableitung  $-10$ .

Wenn man Richtungsableitungen immer so ausrechnen müßte, wie wir das im vorangehenden Beispiel gemacht haben, dann wäre das mühsam. Glücklicherweise kann man es sich einfacher machen.

## 7.1.3 Partielle Ableitungen

Es gibt spezielle Richtungsableitungen, die besonders wichtig sind. Das sind die Richtungsableitungen, deren Richtungsvektoren parallel zu den Koordinatenachsen sind.

(7.3) BEISPIEL:

Wir setzen das vorangegangene Beispiel fort mit anderen Richtungsvektoren. Als Fußpunkt verwenden wir  $\mathbf{x}_0 = (a, b)^t$ .

Es sei  $\mathbf{v} = (1, 0)^t$ . Die Gerade, entlang der wir die Funktion  $f$  nun untersuchen müssen, lautet

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a+t \\ b \end{pmatrix},$$

also

$$\begin{aligned} x_1 &= a + t \\ x_2 &= b \end{aligned}$$

Wir setzen diese Gerade nun in die Funktion  $f$  ein und erhalten so

$$g(t) = f(a+t, b) = 3(a+t)^2 - b^2$$

Daraus folgt  $g'(0) = 6a$ .

Dieses Ergebnis würden wir aber auch erhalten, wenn wir die Funktion  $f$  einfach nach der ersten Variablen differenzieren würden (die zweite Variable dabei konstant haltend). Man schreibt dann

$$\frac{\partial}{\partial x_1} f(x_1, x_2) = \frac{\partial}{\partial x_1} (3x_1^2 - x_2^2) = 6x_1$$

und spricht von der **partiellen Ableitung** von  $f$  nach  $x_1$ .

Es sei  $\mathbf{v} = (0, 1)^t$ . Die Gerade, entlang der wir die Funktion  $f$  nun untersuchen müssen, lautet

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ b+t \end{pmatrix},$$

also

$$\begin{aligned} x_1 &= a \\ x_2 &= b + t \end{aligned}$$

Wir setzen diese Gerade nun in die Funktion  $f$  ein und erhalten so

$$g(t) = f(a, b+t) = 3a^2 - (b+t)^2$$

Daraus folgt  $g'(0) = -2b$ .

Dieses Ergebnis würden wir aber auch erhalten, wenn wir die Funktion  $f$  einfach nach der zweiten Variablen differenzieren würden (die erste Variable dabei konstant haltend). Man schreibt dann

$$\frac{\partial}{\partial x_2} f(x_1, x_2) = \frac{\partial}{\partial x_2} (3x_1^2 - x_2^2) = -2x_2$$

und spricht von der **partiellen Ableitung** von  $f$  nach  $x_2$ .

---

Dies führt uns zur allgemeinen Definition von partiellen Ableitungen.

(7.4) DEFINITION:

Es sei  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , eine Funktion. Unter einer **partiellen Ableitung** versteht man die Ableitung der Funktion nach einer einzigen Variablen, wobei die anderen Variablen als Konstante behandelt werden.

Die partielle Ableitung von  $f$  nach  $x_i$  an der Stelle  $\mathbf{x}_0$  bezeichnet man durch

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0)$$

In unserem Beispiel haben wir plausibel gemacht, dass partielle Ableitungen nichts anderes sind als Richtungsableitungen entlang der Koordinatenachsen.

(7.5) BEISPIEL:

1. Man berechne die partiellen Ableitungen der Funktion  $f(x_1, x_2) = 3x_1^2 - 2x_1x_2 + 5x_2$  an der Stelle  $\mathbf{x}_0 = (3, -1)^t$ .

Lösung: Es ist

$$\frac{\partial f}{\partial x_1} = 6x_1 - 2x_2, \quad \frac{\partial f}{\partial x_2} = -2x_1 + 5.$$

Daraus folgt

$$\frac{\partial f}{\partial x_1}(3, -1) = 20, \quad \frac{\partial f}{\partial x_2}(3, -1) = -1.$$

2. Man berechne die partiellen Ableitungen der Funktion  $f(x_1, x_2) = x_1^2 e^{-3x_2^2}$  an der Stelle  $\mathbf{x}_0 = (1, 0)^t$ .

Lösung: Es ist

$$\frac{\partial f}{\partial x_1} = 2x_1 e^{-3x_2^2}, \quad \frac{\partial f}{\partial x_2} = -12x_1^2 x_2 e^{-3x_2^2}.$$

Daraus folgt

$$\frac{\partial f}{\partial x_1}(1, 0) = 2, \quad \frac{\partial f}{\partial x_2}(1, 0) = 0.$$

## 7.1.4 Der Gradient

Ein besondere Rolle spielt der Vektor, den man aus allen partiellen Ableitungen bilden kann.

(7.6) DEFINITION:

Es sei  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , eine Funktion. Unter dem **Gradienten** der Funktion an der Stelle  $\mathbf{x}_0$  versteht man den Spaltenvektor

$$\nabla f(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}$$

(7.7) BEISPIEL:

In den folgenden Beispielen berechnen wir die Gradienten von linearen und quadratischen Funktionen mit mehreren Variablen. Es zeigt sich dabei, dass die entstehenden Formeln eine grosse Ähnlichkeit zu den uns schon bekannten Ableitungsformeln besitzen.

1. Man berechne den Gradienten der linearen Funktion  $f(\mathbf{x}) = \mathbf{a}^t \mathbf{x}$ .

Lösung: Es ist

$$f(\mathbf{x}) = f(x_1, x_2) = b_1 x_1 + b_2 x_2$$

Daraus ergeben sich die partiellen Ableitungen

$$\frac{\partial f}{\partial x_1} = b_1, \quad \frac{\partial f}{\partial x_2} = b_2$$

Daher lautet der Gradient:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{b}$$

Der Gradient einer linearen Funktion ist also konstant, und ist gleich dem Koeffizientenvektor der linearen Funktion.

2. Man berechne den Gradienten der quadratischen Funktion  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{x}$ .

Lösung: Es ist

$$f(\mathbf{x}) = f(x_1, x_2) = x_1^2 + x_2^2$$

Daraus ergeben sich die partiellen Ableitungen

$$\frac{\partial f}{\partial x_1} = 2x_1, \quad \frac{\partial f}{\partial x_2} = 2x_2$$

Daher lautet der Gradient:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = 2\mathbf{x}$$

3. Man berechne den Gradienten der quadratischen Funktion  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$ .

Lösung: Es ist

$$f(\mathbf{x}) = f(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

Daraus ergeben sich die partiellen Ableitungen

$$\frac{\partial f}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2, \quad \frac{\partial f}{\partial x_2} = 2a_{12}x_1 + 2a_{22}x_2$$

Daher lautet der Gradient:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2a_{11}x_1 + 2a_{12}x_2 \\ 2a_{12}x_1 + 2a_{22}x_2 \end{pmatrix} = 2\mathbf{A}\mathbf{x}$$

4. Man berechne den Gradienten der quadratischen Funktion  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{b}^t \mathbf{x}$ .

Lösung: Es ist

$$f(\mathbf{x}) = f(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + b_1x_1 + b_2x_2$$

Daraus ergeben sich die partiellen Ableitungen

$$\frac{\partial f}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2 + b_1, \quad \frac{\partial f}{\partial x_2} = 2a_{12}x_1 + 2a_{22}x_2 + b_2$$

Daher lautet der Gradient:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2a_{11}x_1 + 2a_{12}x_2 + b_1 \\ 2a_{12}x_1 + 2a_{22}x_2 + b_2 \end{pmatrix} = 2\mathbf{A}\mathbf{x} + \mathbf{b}$$

Der Gradient ist von zweierlei Bedeutung:

1. Mit Hilfe des Gradienten kann man Richtungsableitungen sehr einfach ausrechnen.
2. Der Gradient zeigt in die Richtung des stärksten Anstiegs des Funktionsgraphen.

Wir formulieren diese beiden wichtigen Tatsachen als Lehrsätze und begründen sie auch.

(7.8) SATZ:

Es sei  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , eine Funktion. Die Richtungsableitung von  $f$  an der Stelle  $\mathbf{x}_0$  in Richtung  $\mathbf{v}$  beträgt

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{v} = \frac{\partial f}{\partial x_1}(\mathbf{x}) \cdot v_1 + \frac{\partial f}{\partial x_2}(\mathbf{x}) \cdot v_2$$

Sie ist also gleich dem Skalarprodukt des Gradienten mit dem Richtungsvektor.

BEGRÜNDUNG: Wir müssen nachweisen, dass

$$\left. \frac{d}{dt} f(\mathbf{x} + t\mathbf{v}) \right|_{t=0} = \nabla f(\mathbf{x}) \cdot \mathbf{v}$$

Wir tun dies nicht in voller Allgemeinheit, sondern nur für quadratische Funktionen der Form  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$ .

Es ist

$$\begin{aligned} f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) &= (\mathbf{x} + t\mathbf{v})^t \mathbf{A} (\mathbf{x} + t\mathbf{v}) - \mathbf{x}^t \mathbf{A} \mathbf{x} \\ &= 2t\mathbf{x}^t \mathbf{A} \mathbf{v} + t^2 \mathbf{v}^t \mathbf{A} \mathbf{v} \\ &= t \nabla f(\mathbf{x}) \cdot \mathbf{v} + t^2 \mathbf{v}^t \mathbf{A} \mathbf{v} \end{aligned}$$

Daraus folgt

$$\left. \frac{d}{dt} f(\mathbf{x} + t\mathbf{v}) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} = \lim_{t \rightarrow 0} \nabla f(\mathbf{x}) \cdot \mathbf{v} + t \mathbf{v}^t \mathbf{A} \mathbf{v} = \nabla f(\mathbf{x}) \cdot \mathbf{v}$$

□

(7.9) SATZ:

Es sei  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , eine Funktion. Dann ist die Richtungsableitung an jeder Stelle in jener Richtung maximal, die mit der Richtung des Gradienten übereinstimmt: **Der Gradient gibt die Richtung des stärksten Anstiegs an.** Das heisst: Für alle Richtungsvektoren  $\mathbf{v}$  mit  $\|\mathbf{v}\| = 1$  gilt

$$\nabla f(\mathbf{x}) \cdot \mathbf{v} \leq \nabla f(\mathbf{x}) \cdot \mathbf{v}_{\max} \quad \text{für} \quad \mathbf{v}_{\max} = \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$$

BEGRÜNDUNG: Die behauptete Ungleichung folgt unmittelbar aus der uns wohlbekannten Ungleichung von Cauchy-Schwarz:

$$\nabla f(\mathbf{x}) \cdot \mathbf{v} \leq \|\nabla f(\mathbf{x})\| \|\mathbf{v}\| = \|\nabla f(\mathbf{x})\| = \nabla f(\mathbf{x}) \cdot \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$$

□

## Gradientenfelder

Die geometrische Bedeutung des Gradienten besteht also darin, dass er als Pfeil gezeichnet stets in die Richtung des stärksten Anstiegs des Funktionsgraphen zeigt. Deshalb sind Gradienten eine wichtige Ergänzung eines Niveauliniendiagramms.

In unserer Library **LinAlg** stellen wir die Funktion **gfield** (Gradientenfeld) zu Verfügung, die ein Niveauliniendiagramm mit Gradienten zeichnet.

Das Beispiel

```
gfield( from=-3, to=3, "x^2+y^2" )
```

zeigt die Gradienten eines nach oben offenen Paraboloids, bei dem die Gradienten nach außen weisen. Andere Beispiele sind ebenfalls sehr illustrativ. Wir überlassen sie dem Leser.

### 7.1.5 Aufgaben

#### (7.10) TESTFRAGEN

1. Was versteht man unter der Richtungsableitung ? Erklären Sie den Begriff anschaulich und formal.
2. Was versteht man unter einer partiellen Ableitung ? Erklären Sie den Begriff anschaulich und formal.
3. Was ist der Gradient ? Was ist seine anschauliche Bedeutung ?
4. Wie kann man Richtungsableitungen mit Hilfe des Gradienten berechnen ?
5. Wie berechnet man den Gradienten einer quadratischen Funktion ? Geben Sie die Matrixdarstellung an.

## 7.2 Globale Optimierung

### 7.2.1 Kritische Punkte

Wenn wir Maximalstellen oder Minimalstellen einer Funktion von mehreren Variablen finden wollen, suchen wir als Kandidaten solcher Extremstellen zunächst sogenannte kritische Punkte.

(7.11) DEFINITION:

*Unter einem kritischen Punkt einer Funktion  $y = f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ , versteht man einen Punkt  $\mathbf{x}_0$ , an dem der Gradient gleich Null ist, dh. wo  $\nabla f(\mathbf{x}_0) = 0$  ist.*

Mit anderen Worten: An einem kritischen Punkt sind alle partiellen Ableitungen gleich Null.

Die Bedeutung von kritischen Punkten für die Optimierungstheorie ergibt sich durch den folgenden Lehrsatz.

(7.12) SATZ:

*Jede Maximalstelle oder Minimalstelle einer differenzierbaren Funktion, die sich im Inneren ihres Definitionsbereichs befindet, ist ein kritischer Punkt, dh. der Gradient ist dort gleich Null.*

BEGRÜNDUNG: Sehen wir uns die Begründung dafür anhand einer Maximalstelle an. Es sei  $\mathbf{x}_0$  eine Maximalstelle, dh.

$$f(\mathbf{x}_0) \geq f(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n.$$

Diese Extremaleigenschaft gilt dann insbesondere auch entlang jeder Geraden. Wenn also  $\mathbf{v}$  ein beliebiger Richtungsvektor ist, dann gilt

$$f(\mathbf{x}_0) \geq f(\mathbf{x}_0 + t\mathbf{v}) \quad \text{für alle } t \in \mathbb{R}.$$

Für die Funktion  $g(t) := f(\mathbf{x}_0 + t\mathbf{v})$  gilt daher  $g(0) \geq g(t)$  für alle  $t \in \mathbb{R}$ : Sie besitzt ein Maximum an der Stelle  $t = 0$ . Daher muss ihre Ableitung dort gleich Null sein:  $g'(0) = 0$ . Diese Ableitung ist aber nichts anderes als die Richtungsableitung der Funktion  $f$  an der Stelle  $\mathbf{x}_0$  in Richtung  $\mathbf{v}$ .

Wir haben damit gesehen, dass an einer Maximalstelle alle Richtungsableitungen gleich Null sein müssen. Insbesondere müssen daher die partiellen Ableitungen gleich Null sein.  $\square$

Wenn auch jede Maximalstelle oder Minimalstelle (im Innern des Definitionsbereichs) ein kritischer Punkt sein muss, müssen dennoch nicht alle kritischen Punkte auch Maximalstelle oder Minimalstellen sein.

Bei konvexen und konkaven Funktionen sind jedoch die Verhältnisse sehr einfach. Der folgende Satz sagt uns, dass bei konvexen und konkaven Funktionen mit der Bestimmung der kritischen Punkte das Optimierungsproblem bereits gelöst ist.

(7.13) SATZ:

*Jeder kritische Punkt einer konvexen Funktion ist zugleich ein globales Minimum.  
Jeder kritische Punkt einer konkaven Funktion ist zugleich ein globales Maximum.*

BEGRÜNDUNG: Es sei  $y = f(\mathbf{x})$  eine konvexe Funktion und  $\mathbf{x}_0$  sei ein kritischer Punkt. Wir wollen zeigen, dass dann  $\mathbf{x}_0$  auch ein globales Minimum ist, dh. dass

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n.$$

Es sei also  $\mathbf{x}$  ein beliebiger weiterer Punkt. Wir müssen nachweisen, dass  $f(\mathbf{x}_0) \leq f(\mathbf{x})$ .

Wir betrachten nun  $f$  entlang der Geraden von  $\mathbf{x}_0$  nach  $\mathbf{x}$ , dh. wir betrachten die Funktion

$$g(t) := f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)), \quad t \in \mathbb{R}.$$

Da die Funktion  $f$  insgesamt eine konvexe Funktion ist, ist natürlich auch ihre Einschränkung auf die Gerade von  $\mathbf{x}_0$  nach  $\mathbf{x}$  eine konvexe Funktion. Daher ist  $g(t)$  eine konvexe Funktion.

Da  $g(t)$  eine konvexe Funktion ist, ist  $g''(t) \geq 0$  und daher  $g'(t)$  monoton wachsend. Nun ist aber  $\mathbf{x}_0$  ein kritischer Punkt. Daher ist  $g'(0) = 0$ . Wegen der Monotonie von  $g'$  ist nun  $g'(t) \geq 0$  für  $t \geq 0$ , und somit  $g(t)$  ebenfalls monoton wachsend für  $t \geq 0$ . Daraus folgt

$$g(0) \leq g(1), \quad \text{dh.} \quad f(\mathbf{x}_0) \leq f(\mathbf{x}).$$

 $\square$

## 7.2.2 Homogene quadratische Funktionen

Für homogene quadratische Funktionen ist die Bestimmung von kritischen Punkten ziemlich einfach.

Es sei  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$  eine homogene quadratische Funktion. Der Gradient lautet dann  $\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x}$ . Daher ist zunächst  $\mathbf{x} = 0$  sicher ein kritischer Punkt. Ob es noch weitere kritische Punkte gibt, hängt davon ab, ob das lineare Gleichungssystem

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} = 0$$

abgesehen von  $\mathbf{x} = 0$  noch weitere Lösungen besitzt. Das wiederum hängt mit der Invertierbarkeit der Matrix  $\mathbf{A}$  zusammen.

(7.14) SATZ:

*Ein symmetrische Matrix  $\mathbf{A}$  ist genau dann invertierbar, wenn alle Eigenwerte von Null verschieden sind.*

BEGRÜNDUNG: Die Matrix  $\mathbf{A}$  ist, wie wir wissen, genau dann invertierbar, wenn das Gleichungssystem  $\mathbf{A}\mathbf{x} = \mathbf{b}$  stets eindeutig lösbar ist. Ist  $\mathbf{T}^t \mathbf{A} \mathbf{T} = \mathbf{D}$  die Diagonalisierung der Matrix  $\mathbf{A}$ , und bezeichnen wir  $\mathbf{x} = \mathbf{T}\mathbf{y}$  und  $\mathbf{b} = \mathbf{T}\mathbf{c}$ , dann ist

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{A}\mathbf{T}\mathbf{y} = \mathbf{T}\mathbf{c} \quad \Leftrightarrow \quad \mathbf{T}^t \mathbf{A} \mathbf{T} \mathbf{y} = \mathbf{c} \quad \Leftrightarrow \quad \mathbf{D}\mathbf{y} = \mathbf{c}.$$

Das Gleichungssystem links ist also genau dann eindeutig lösbar, wenn das Gleichungssystem mit der Diagonalmatrix  $\mathbf{D}$  eindeutig lösbar ist. Das ist aber nur dann der Fall, wenn die Diagonalmatrix  $\mathbf{D}$  vollen Rang hat, dh. wenn alle Diagonalelemente (Eigenwerte) von Null verschieden sind.  $\square$

Damit können wir zusammenfassen:

(7.15) SATZ:

*Es sei  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$  eine homogene quadratische Funktion.*

- 1. Wenn alle Eigenwerte von Null verschieden sind, dann ist  $\mathbf{x} = 0$  der einzige kritische Punkt.*
- 2. Wenn alle Eigenwerte positiv sind, dann ist die Funktion konvex, und  $\mathbf{x} = 0$  ist das einzige globale Minimum.*
- 3. Wenn alle Eigenwerte negativ sind, dann ist die Funktion konkav, und  $\mathbf{x} = 0$  ist das einzige globale Maximum.*

Die Optimierung von homogenen quadratischen Funktionen ist offenbar ein triviale Aufgabe, und deshalb auch von geringer praktischer Bedeutung.

## 7.2.3 Inhomogene quadratische Funktionen

Bei inhomogenen quadratischen Funktionen sieht die Sache bereits ganz anders aus.

Es sei

$$f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{b}^t \mathbf{x} + c$$

eine (inhomogene) quadratische Funktion. Der Gradient lautet

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + \mathbf{b}$$

Um kritische Punkte zu finden, müssen wir das Gleichungssystem

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + \mathbf{b} = 0, \quad \text{dh.} \quad \mathbf{A}\mathbf{x} = -\frac{1}{2}\mathbf{b}$$

lösen. Wenn die Matrix  $\mathbf{A}$  invertierbar ist, dann ist dieses Gleichungssystem eindeutig lösbar und

$$\mathbf{x}_0 = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$$

ist der einzige kritische Punkt der Funktion  $f$ . Wenn aber die Matrix  $\mathbf{A}$  nicht invertierbar ist, dann sind beide weitere Lösungsfälle möglich: Das Gleichungssystem kann unendlich viele Lösungen besitzen, oder es kann unlösbar sein.

Wenn wir kritische Punkt gefunden haben, dann hängt ihre Beurteilung wieder davon ab, ob die Funktion  $f$  konvex oder konkav ist.

Damit können wir wieder zusammenfassen:

(7.16) SATZ:

*Es sei  $f(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x} + \mathbf{b}^t\mathbf{x} + c$  eine inhomogene quadratische Funktion.*

*1. Wenn alle Eigenwerte der Matrix  $\mathbf{A}$  von Null verschieden sind, dann ist  $\mathbf{x}_0 = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$  der einzige kritische Punkt.*

*2. Wenn alle Eigenwerte der Matrix  $\mathbf{A}$  positiv sind, dann ist die Funktion konvex, und  $\mathbf{x}_0$  ist das einzige globale Minimum.*

*3. Wenn alle Eigenwerte der Matrix  $\mathbf{A}$  negativ sind, dann ist die Funktion konkav, und  $\mathbf{x}_0$  ist das einzige globale Maximum.*

(7.17) BEISPIEL:

Gegeben ist die quadratische Funktion

$$f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + 3x_2^2 - x_1 + 4x_2 + 10$$

Man bestimme kritische Punkte und beurteile, ob es sich um eine Maximumstelle oder eine Minimumstelle handelt.

Lösung: Wir stellen die Funktion zunächst in Matrixform dar. Es ist

$$f(x_1, x_2) = (x_1, x_2) \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (-1, 4) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 10$$

Mit

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}, \quad c = 10$$

gilt daher

$$f(\mathbf{x}) = \mathbf{x}^t\mathbf{A}\mathbf{x} + \mathbf{b}^t\mathbf{x} + c$$

Da  $\mathbf{A}$  eine reguläre Matrix ist, gibt es nur einen kritischen Punkt, nämlich

$$\mathbf{x}_0 = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$$

Wir berechnen ihn durch

```

> A=cbind(c(2,1),c(1,3))
> b=cbind(c(-1,4))
> x.krit=-solve(A)%*%b/2
> x.krit
      [,1]
[1,]  0.7
[2,] -0.9

```

Es ist also

$$\mathbf{x}_0 = \begin{pmatrix} 0.7 \\ -0.9 \end{pmatrix}$$

Aus

```

> eigen(A)$values
[1] 3.618034 1.381966

```

folgt, dass die Funktion  $f$  konvex ist. Daher ist der kritische Punkt  $\mathbf{x}_0$  eine globale Minimumstelle. Das Minimum selbst beträgt

```

> t(x.krit)%*%A%*%x.krit+t(b)%*%x.krit+10
      [,1]
[1,] 7.85

```

Wir zeichnen das Niveauliniendiagramm auf einem Bereich, in dem der kritische Punkt im Zentrum liegt:

```

> x=seq(from=-3,to=4,by=0.5)
> y=seq(from=-4,to=3,by=0.5)
> ff=function(x,y) 2*x^2+2*x+y+3*y^2-x+4*y+10
> z=outer(x,y,ff)
> contour(x,y,z)

```

Wir sehen, dass der kritische Punkt tatsächlich im Zentrum der ansteigenden Niveaulinien liegt.

## 7.2.4 Aufgaben

(7.18) **Aufgabe:** Für die folgenden quadratischen Funktionen beantworte man die Fragen:

1. Wie lautet die Matrixdarstellung ?
2. Bestimmen Sie kritische Punkte der Funktion.
3. Beurteilen Sie die kritischen Punkte hinsichtlich der Frage, ob es sich um Maximumstellen oder um Minimumstellen handelt.

Angaben:

1.  $f(x_1, x_2) = -4x_1^2 - 4x_1x_2 - x_2^2 + 2x_1 - x_2 - 3$
2.  $f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + 2x_2^2 + 2x_1 - 3x_2 + 5$
3.  $f(x_1, x_2) = 3x_1^2 - 4x_1x_2 + 2x_2^2 - 4x_1 - x_2 + 1$
4.  $f(x_1, x_2) = 2x_1^2 + 8x_1x_2 + 4x_2^2 - x_1 - x_2$

5.  $f(x_1, x_2) = -4x_1^2 - 6x_1x_2 - 2x_2^2 + 5x_1 + 3x_2 - 8$
6.  $f(x_1, x_2) = -3x_1^2 - 4x_1x_2 + 2x_1 - x_2 - 3$
7.  $f(x_1, x_2) = 3x_1^2 + 8x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 + 5$
8.  $f(x_1, x_2) = -3x_1^2 + 10x_1x_2 + 1x_2^2 - 4x_1 - x_2 + 1$
9.  $f(x_1, x_2) = -4x_1^2 - 4x_1x_2 - 2x_2^2 - x_1 - x_2$
10.  $f(x_1, x_2) = -2x_1^2 + 4x_1x_2 - 4x_2^2 + 5x_1 + 3x_2 - 8$
11.  $f(x_1, x_2) = 3x_1^2 + 2x_1x_2 + 4x_2^2 + 2x_1 - x_2 - 3$
12.  $f(x_1, x_2) = 5x_1^2 + 6x_1x_2 + 4x_2^2 + 2x_1 - 3x_2 + 5$
13.  $f(x_1, x_2) = 5x_1^2 - 6x_1x_2 + 4x_2^2 - 4x_1 - x_2 + 1$
14.  $f(x_1, x_2) = -2x_1^2 + 2x_1x_2 - 2x_2^2 - x_1 - x_2$
15.  $f(x_1, x_2) = -2x_1^2 + 2x_1x_2 - 4x_2^2 + 5x_1 + 3x_2 - 8$
16.  $f(x_1, x_2) = 3x_1^2 + 6x_1x_2 + 3x_2^2 + 2x_1 - x_2 - 3$
17.  $f(x_1, x_2) = 4x_1^2 + 4x_1x_2 + x_2^2 + 2x_1 - 3x_2 + 5$
18.  $f(x_1, x_2) = -x_1^2 + 2x_1x_2 - x_2^2 - 4x_1 - x_2 + 1$
19.  $f(x_1, x_2) = -4x_1^2 + 2x_1x_2 - 3x_2^2 - x_1 - x_2$
20.  $f(x_1, x_2) = -4x_1^2 + 4x_1x_2 - x_2^2 + 5x_1 + 3x_2 - 8$

#### (7.19) TESTFRAGEN

1. Was ist ein kritischer Punkt einer Funktion mit mehreren Variablen ?
2. Worin besteht die Bedeutung von kritischen Punkten für Optimierungsprobleme ?
3. Unter welchen Voraussetzungen kann man kritische Punkte ohne weitere Untersuchung als globale Extremstellen interpretieren ?
4. Was sind die kritischen Punkte einer homogenen quadratischen Funktion ? Wie beurteilt man sie hinsichtlich ihrer Extremaleigenschaften ?
5. Wie berechnet man die kritischen Punkte einer inhomogenen quadratischen Funktion ? Wie beurteilt man sie hinsichtlich ihrer Extremaleigenschaften ?

## 7.3 Optimierung mit Nebenbedingungen

In praktischen Anwendungen treten häufig quadratische Optimierungsprobleme auf. Allerdings sind diese meistens nicht so einfach, wie im vorangehenden Abschnitt, wo ein Punkt  $\mathbf{x}_0$  gesucht wird, an dem die Funktion  $y = f(\mathbf{x})$  ihr Maximum oder ihr Minimum annimmt. Es ist vielmehr so, dass bei der Optimierung zusätzliche Nebenbedingungen erfüllt werden müssen.

Ein wichtiges Beispiel aus dem Bereich der Wirtschaftswissenschaften ist die Bestimmung eines optimalen Aktienportfolios nach Markowitz. Wir behandeln dieses Beispiel im letzten Abschnitt dieses Kapitels.

Das Problem der Optimierung unter Nebenbedingungen, die durch Gleichungen beschrieben werden, läßt sich mathematisch unter sehr allgemeinen Voraussetzungen behandeln.

Es gibt dafür ein universelles Lösungsverfahren, die sogenannte **Methode der Lagrange-Multiplikatoren**. Wir werden diese Methode nur für den Fall der Optimierung von quadratischen Funktionen unter linearen Nebenbedingungen vorstellen. Dies reicht aus, um das Problem der Portfolio-Optimierung lösen zu können.

### 7.3.1 Optimierung unter einer linearen Nebenbedingung

Es sei  $y = f(\mathbf{x})$  eine Funktion. Unsere Aufgabe bestehe darin, einen Punkt  $\mathbf{x}_0$  zu finden, der erstens eine bestimmte **Nebenbedingung** erfüllt, und der zweitens unter allen weiteren Punkten, die die gleiche Nebenbedingung erfüllen, optimal ist, dh. eine Maximumstelle bzw. eine Minimumstelle ist.

Sehen wir uns ein Beispiel an.

---

(7.20) BEISPIEL:

Wir betrachten die Funktion  $f(x_1, x_2) = x_1^2 + x_2^2$  und die Nebenbedingung  $x_1 + x_2 = 1$ . Wir stellen uns die Aufgabe:

Minimiere  $x_1^2 + x_2^2$  unter der Nebenbedingung  $x_1 + x_2 = 1$  !

Natürlich ist diese Aufgabe auf direktem Weg ganz einfach lösbar. Man braucht nur  $x_2$  aus der Nebenbedingung auszurechnen und in die zu optimierende Funktionen einzusetzen:

Aus  $x_1 + x_2 = 1$  folgt  $x_2 = 1 - x_1$  und daher ist

$$x_1^2 + x_2^2 = x_1^2 + (1 - x_1)^2 = 2x_1^2 - 2x_1 + 1$$

Das Minimum finden wir, indem wir die Ableitung gleich Null setzen:

$$4x_1 - 2 = 0 \quad \implies \quad x_1 = \frac{1}{2}, \quad x_2 = \frac{1}{2}.$$

Der gesuchte Minimalpunkt lautet also  $\mathbf{x}_0 = (\frac{1}{2}, \frac{1}{2})$ .

Eine Anmerkung ist noch angebracht, die sich für das weitere als bedeutsam herausstellen wird: Der Minimalpunkt ist keineswegs ein kritischer Punkt der Funktion  $f$  ! Der Gradient von  $f$  an der Stelle  $\mathbf{x}_0$  ist nicht gleich Null, sondern beträgt vielmehr

$$\nabla f(\mathbf{x}_0) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Der Gradient muss auch nicht gleich Null sein, denn wir haben ja keine Minimumstelle schlechthin gesucht, sondern unter einer Nebenbedingung. Interessant ist, dass diese Nebenbedingung im Wert des Gradienten sichtbar ist: Der Gradient ist identisch mit dem Koeffizientenvektor der Nebenbedingung !

---

Im Prinzip kann man Optimierung unter einer linearen Nebenbedingung immer so durchführen, wie wir das in dem vorangehenden Beispiel gemacht haben. Aber man tut das nicht. Man modifiziert vielmehr den Begriff des kritischen Punktes so, dass die Nebenbedingung dabei berücksichtigt wird, und macht sich dann auf die Suche nach solchen kritischen Punkten.

Die Grundlage dafür ist der folgende Lehrsatz.

(7.21) SATZ:

*Es sei  $y = f(\mathbf{x})$  eine Funktion und  $\mathbf{k} \cdot \mathbf{x} = c$  sei eine lineare Nebenbedingung. Wenn  $\mathbf{x}_0$  entweder eine Minimalstelle von  $f$  unter dieser Nebenbedingung ist, dh.*

$$f(\mathbf{x}_0) = \text{Min}\{f(\mathbf{x}) : \mathbf{k} \cdot \mathbf{x} = c\},$$

*oder eine Maximalstelle von  $f$  unter dieser Nebenbedingung ist, dh.*

$$f(\mathbf{x}_0) = \text{Max}\{f(\mathbf{x}) : \mathbf{k} \cdot \mathbf{x} = c\},$$

*dann ist der Gradient  $\nabla f(\mathbf{x}_0)$  proportional zu  $\mathbf{k}$ , dh. es gibt eine Zahl  $\lambda$ , sodass  $\nabla f(\mathbf{x}_0) = \lambda \mathbf{k}$ .*

BEGRÜNDUNG: Wenn

$$f(\mathbf{x}_0) = \text{Min}\{f(\mathbf{x}) : \mathbf{k} \cdot \mathbf{x} = c\},$$

dann ist  $\mathbf{x}_0$  auch Minimumstelle entlang aller Geraden, die zur Gänze die Nebenbedingung erfüllen. Wenn die Gerade  $\mathbf{x} = \mathbf{x}_0 + t\mathbf{h}$  für alle  $t \in \mathbb{R}$  die Nebenbedingung  $\mathbf{k} \cdot \mathbf{x} = c$  erfüllt, dann ergibt sich für  $t = 1$

$$\mathbf{k} \cdot \mathbf{x} = \mathbf{k} \cdot \mathbf{x}_0 + \mathbf{k} \cdot \mathbf{h} \quad \Longrightarrow \quad c = c + \mathbf{k} \cdot \mathbf{h} \quad \Longrightarrow \quad \mathbf{k} \cdot \mathbf{h} = 0.$$

Die Richtungsvektoren von Geraden, die die Nebenbedingung erfüllen, sind daher genau jene Vektoren  $\mathbf{h}$ , für die  $\mathbf{k} \cdot \mathbf{h} = 0$ .

Wenn  $\mathbf{x}_0$  eine Minimumstelle von  $f$  entlang solcher Geraden ist, muss die Richtungsableitung gleich Null sein, also  $\nabla f(\mathbf{x}_0) \cdot \mathbf{h} = 0$ . Damit haben wir folgende Aussage erreicht:

$$\mathbf{k} \cdot \mathbf{h} = 0 \quad \Longrightarrow \quad \nabla f(\mathbf{x}_0) \cdot \mathbf{h} = 0$$

Aus Lemma 7.28 folgt nun, dass  $\nabla f(\mathbf{x}_0)$  proportional zu  $\mathbf{k}$  ist. □

So kommen wir nun zur Definition von kritischen Punkten für Optimierungsprobleme mit Nebenbedingungen.

(7.22) DEFINITION:

*Unter einem kritischen Punkt der Funktion  $y = f(\mathbf{x})$  unter der Nebenbedingung  $\mathbf{k} \cdot \mathbf{x} = c$  versteht man einen Punkt  $\mathbf{x}_0$ , für den der Gradient  $\nabla f(\mathbf{x}_0)$  proportional zu  $\mathbf{k}$  ist, dh. für den es eine Zahl  $\lambda$  gibt, sodass  $\nabla f(\mathbf{x}_0) = \lambda \mathbf{k}$ .*

Alle Extremstellen unter einer linearen Nebenbedingung sind kritischen Punkte gemäß dieser Definition. Natürlich sind nicht alle kritischen Punkte tatsächlich auch Extremstellen. Aber wenn die Funktion  $f$ , die optimiert werden soll, konvex oder konkav ist, dann ändert sich diese Eigenschaft auch unter der **linearen** Nebenbedingung nicht, und wir können in diesem Fall kritische Punkte ohne weitere Analyse als Maxima oder Minima interpretieren.

## Die Methode von Lagrange

---

Die rechnerische Durchführung der Bestimmung von kritischen Punkten verläuft aus historischen Gründen nach einem bestimmten Schema ab, das man als **Methode der Lagrange-Multiplikatoren** bezeichnet.

Eigentlich müsste man nur Punkte  $\mathbf{x}$  suchen, für die es eine Zahl  $\lambda$  gibt, sodass

$$\nabla f(\mathbf{x}) + \lambda \mathbf{k} = 0$$

In der Praxis geht man aber immer so vor: Man definiert die sogenannte **Lagrange-Funktion**

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda(\mathbf{k} \cdot \mathbf{x} - c)$$

Von dieser Funktion sucht man nun kritische Punkte im ursprünglichen Sinn, dh. man bildet die partiellen Ableitungen nach  $\mathbf{x}$  und  $\lambda$  und setzt diese gleich Null. Das führt genau auf die Gleichungen, die wir lösen müssen, denn

$$\begin{aligned} \frac{\partial L}{\partial x_1}(\mathbf{x}, \lambda) &= \frac{\partial f}{\partial x_1} + \lambda k_1 = 0 \\ \frac{\partial L}{\partial x_2}(\mathbf{x}, \lambda) &= \frac{\partial f}{\partial x_2} + \lambda k_2 = 0 \\ &\vdots \\ \frac{\partial L}{\partial x_n}(\mathbf{x}, \lambda) &= \frac{\partial f}{\partial x_n} + \lambda k_n = 0 \\ \frac{\partial L}{\partial \lambda}(\mathbf{x}, \lambda) &= \mathbf{k} \cdot \mathbf{x} - c = 0 \end{aligned}$$

(7.23) BEISPIEL:

Wir betrachten nochmals die Funktion  $f(x_1, x_2) = x_1^2 + x_2^2$  und die Nebenbedingung  $x_1 + x_2 = 1$ . Wir stellen uns wieder die Aufgabe:

Minimiere  $x_1^2 + x_2^2$  unter der Nebenbedingung  $x_1 + x_2 = 1$  !

Wir lösen die Aufgabe aber jetzt mit der Methode von Lagrange.

Die Lagrangefunktion lautet:

$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1)$$

Die Bildung der partiellen Ableitungen ergibt:

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 2x_1 + \lambda = 0 \\ \frac{\partial L}{\partial x_2} &= 2x_2 + \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= x_1 + x_2 - 1 = 0 \end{aligned}$$

Also müssen  $x_1$  und  $x_2$  gleich gross sein und ihre Summe muss gleich 1 sein. Das führt zur Lösung  $\mathbf{x}_0 = (\frac{1}{2}, \frac{1}{2})$ .

Dieser kritische Punkt ist ein Minimum, denn die Funktion  $f$  ist auf der Geraden, die durch die Nebenbedingung definiert ist, konvex.

Sehen wir uns die allgemeine Struktur des Problems an.

Es sei  $f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}$  eine homogene quadratische Funktion und  $\mathbf{k} \cdot \mathbf{x} = c$  eine lineare Nebenbedingung. Wir wollen kritische Punkte berechnen.

Die Lagrange-Funktion lautet

$$L(\mathbf{x}, \lambda) = \mathbf{x}^t \mathbf{A} \mathbf{x} + \lambda(\mathbf{k} \cdot \mathbf{x} - c)$$

Wir bilden partielle Ableitungen nach  $\mathbf{x}$  und  $\lambda$ :

$$\begin{aligned}\nabla_{\mathbf{x}} L &= 2\mathbf{A}\mathbf{x} + \lambda\mathbf{k} \\ \nabla_{\lambda} L &= \mathbf{k} \cdot \mathbf{x} - c\end{aligned}$$

Um kritische Punkte zu finden, müssen wir also das Gleichungssystem

$$\begin{aligned}2\mathbf{A}\mathbf{x} + \lambda\mathbf{k} &= 0 \\ \mathbf{k}^t \mathbf{x} &= c\end{aligned}$$

mit den  $n+1$  Unbekannten  $x_1, x_2, \dots, x_n, \lambda$  lösen. Jede Lösung dieses Gleichungssystems ist ein kritischer Punkt, sofern sie die Nebenbedingung erfüllt, was man aber immer durch geeignete Wahl von freien Parametern erreichen kann. Das Gleichungssystem hat die Gleichungsmatrix

$$\left( \begin{array}{cccc|c} 2a_{11} & 2a_{12} & \dots & 2a_{1n} & k_1 & 0 \\ 2a_{12} & 2a_{22} & \dots & 2a_{2n} & k_2 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 2a_{1n} & 2a_{2n} & \dots & 2a_{nn} & k_n & 0 \\ k_1 & k_2 & \dots & k_n & 0 & c \end{array} \right)$$

Wenn  $f$  konvex oder konkav ist, können wir den gewonnenen kritischen Punkt sofort als Minimum oder Maximum interpretieren.

(7.24) BEISPIEL:

Man minimiere die Funktion  $f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + x_3^2 - x_1x_2$  unter der Nebenbedingung  $x_2 - x_3 = 3$ .

Lösung: Die Matrix der quadratischen Funktion lautet

```
> A=cbind(c(1,-1,0),c(-1,2,0),c(0,0,1))
> A
      [,1] [,2] [,3]
[1,]    1   -1    0
[2,]   -1    2    0
[3,]    0    0    1
```

Die Eigenwerte

```
> eigen(A)$values
[1] 2.618034 1.000000 0.381966
```

zeigen, dass die quadratische Funktion konvex ist. Die Matrix ist außerdem invertierbar, weil alle Eigenwerte von Null verschieden sind.

Wir bauen die Koeffizientenmatrix des Gleichungssystems auf:

```
> k=c(0,1,-1)
> C=rbind(cbind(2*A,k),c(k,0))
```

```
> C
      k
[1,]  2 -2  0  0
[2,] -2  4  0  1
[3,]  0  0  2 -1
[4,]  0  1 -1  0
```

und anschließend die Gleichungsmatrix

```
> G=cbind(C,c(rep(0,3),3))
> G
      k
[1,]  2 -2  0  0  0
[2,] -2  4  0  1  0
[3,]  0  0  2 -1  0
[4,]  0  1 -1  0  3
```

Das Eliminationsverfahren

```
> rref(G)$echelon.form
      k
[1,]  1  0  0  0  1.5
[2,]  0  1  0  0  1.5
[3,]  0  0  1  0 -1.5
[4,]  0  0  0  1 -3.0
```

liefert uns den kritischen Punkt  $\mathbf{x} = (1.5, 1.5, -1.5)^t$ . Da die Funktion konvex ist, ist dieser kritische Punkt ein Minimum.

## 7.3.2 Optimierung unter mehreren linearen Nebenbedingungen

Es sei wieder  $y = f(\mathbf{x})$  eine Funktion. Unsere Aufgabe bestehe nun darin, einen Punkt  $\mathbf{x}_0$  zu finden, der mehrere **Nebenbedingungen** erfüllt, und der zweitens unter allen weiteren Punkten, die die gleichen Nebenbedingungen erfüllen, optimal ist, dh. eine Maximumstelle bzw. eine Minimumstelle ist.

Die Grundlage dafür ist der folgende Lehrsatz. Wir formulieren ihn für zwei lineare Nebenbedingungen.

(7.25) SATZ:

*Es sei  $y = f(\mathbf{x})$  eine Funktion und  $\mathbf{k}_1 \cdot \mathbf{x} = c_1, \mathbf{k}_2 \cdot \mathbf{x} = c_2$  seien zwei lineare Nebenbedingungen. Wenn  $\mathbf{x}_0$  entweder eine Minimalstelle von  $f$  unter diesen Nebenbedingungen ist, dh.*

$$f(\mathbf{x}_0) = \text{Min}\{f(\mathbf{x}) : \mathbf{k}_1 \cdot \mathbf{x} = c_1, \mathbf{k}_2 \cdot \mathbf{x} = c_2\},$$

*oder eine Maximalstelle von  $f$  unter dieser Nebenbedingung ist, dh.*

$$f(\mathbf{x}_0) = \text{Max}\{f(\mathbf{x}) : \mathbf{k}_1 \cdot \mathbf{x} = c_1, \mathbf{k}_2 \cdot \mathbf{x} = c_2\},$$

*dann gibt es Zahlen  $\lambda_1$  und  $\lambda_2$ , sodass  $\nabla f(\mathbf{x}_0) = \lambda_1 \mathbf{k}_1 + \lambda_2 \mathbf{k}_2$ .*

BEGRÜNDUNG: Wenn

$$f(\mathbf{x}_0) = \text{Min}\{f(\mathbf{x}) : \mathbf{k}_1 \cdot \mathbf{x} = c_1, \mathbf{k}_2 \cdot \mathbf{x} = c_2\},$$

dann ist  $\mathbf{x}_0$  auch Minimumstelle entlang aller Geraden, die zur Gänze beide Nebenbedingungen erfüllen. Wenn die Gerade  $\mathbf{x} = \mathbf{x}_0 + t\mathbf{h}$  für alle  $t \in \mathbb{R}$  beide Nebenbedingungen  $\mathbf{k}_i \cdot \mathbf{x} = c_i$ ,  $i = 1, 2$ , erfüllt, dann ergibt sich für  $t = 1$  und  $i = 1, 2$ , dass

$$\mathbf{k}_i \cdot \mathbf{x} = \mathbf{k}_i \cdot \mathbf{x}_0 + \mathbf{k}_i \cdot \mathbf{h} \quad \Longrightarrow \quad c = c + \mathbf{k}_i \cdot \mathbf{h} \quad \Longrightarrow \quad \mathbf{k}_i \cdot \mathbf{h} = 0.$$

Die Richtungsvektoren von Geraden, die die Nebenbedingung erfüllen, sind daher genau jene Vektoren  $\mathbf{h}$ , für die  $\mathbf{k}_1 \cdot \mathbf{h} = 0$  und  $\mathbf{k}_2 \cdot \mathbf{h} = 0$ .

Wenn  $\mathbf{x}_0$  eine Minimumstelle von  $f$  entlang solcher Geraden ist, muss die Richtungsableitung gleich Null sein, also  $\nabla f(\mathbf{x}_0) \cdot \mathbf{h} = 0$ . Damit haben wir folgende Aussage erreicht:

$$\mathbf{k}_1 \cdot \mathbf{h} = 0, \mathbf{k}_2 \cdot \mathbf{h} = 0 \quad \Longrightarrow \quad \nabla f(\mathbf{x}_0) \cdot \mathbf{h} = 0$$

Aus Lemma 7.29 folgt nun, dass  $\nabla f(\mathbf{x}_0)$  in der Form  $\lambda_1 \mathbf{k}_1 + \lambda_2 \mathbf{k}_2$  dargestellt werden kann.  $\square$

So kommen wir nun zur Definition von kritischen Punkten für Optimierungsprobleme mit Nebenbedingungen.

(7.26) DEFINITION:

Unter einem kritischen Punkt der Funktion  $y = f(\mathbf{x})$  unter zwei Nebenbedingungen  $\mathbf{k}_1 \cdot \mathbf{x} = c_1$  und  $\mathbf{k}_2 \cdot \mathbf{x} = c_2$  versteht man einen Punkt  $\mathbf{x}_0$ , für den es Zahlen  $\lambda_1$  und  $\lambda_2$  gibt, sodass  $\nabla f(\mathbf{x}_0) = \lambda_1 \mathbf{k}_1 + \lambda_2 \mathbf{k}_2$ .

Alle Extremstellen unter linearen Nebenbedingungen sind kritischen Punkte gemäß dieser Definition. Natürlich sind nicht alle kritischen Punkte tatsächlich auch Extremstellen. Aber wenn die Funktion  $f$ , die optimiert werden soll, konvex oder konkav ist, dann ändert sich diese Eigenschaft auch unter **linearen** Nebenbedingungen nicht, und wir können in diesem Fall kritische Punkte ohne weitere Analyse als Maxima oder Minima interpretieren.

## Die Methode von Lagrange

---

Die rechnerische Durchführung der Bestimmung von kritischen Punkten verläuft wieder nach dem Schema ab, das man als **Methode der Lagrange-Multiplikatoren** bezeichnet.

Man definiert die sogenannte **Lagrange-Funktion**

$$L(\mathbf{x}, \lambda_1, \lambda_2) = f(\mathbf{x}) + \lambda_1(\mathbf{k}_1 \cdot \mathbf{x} - c_1) + \lambda_2(\mathbf{k}_2 \cdot \mathbf{x} - c_2)$$

Von dieser Funktion sucht man kritische Punkte im ursprünglichen Sinn, dh. man bildet die partiellen Ableitungen nach  $\mathbf{x}$ ,  $\lambda_1$  und  $\lambda_2$ , und setzt diese gleich Null. Das führt genau auf

die Gleichungen, die wir lösen müssen, denn

$$\begin{aligned}\frac{\partial L}{\partial x_1}(\mathbf{x}, \lambda_1, \lambda_2) &= \frac{\partial f}{\partial x_1} + \lambda_1 k_{11} + \lambda_2 k_{21} = 0 \\ \frac{\partial L}{\partial x_2}(\mathbf{x}, \lambda_1, \lambda_2) &= \frac{\partial f}{\partial x_2} + \lambda_1 k_{12} + \lambda_2 k_{22} = 0 \\ &\vdots \\ \frac{\partial L}{\partial x_n}(\mathbf{x}, \lambda_1, \lambda_2) &= \frac{\partial f}{\partial x_n} + \lambda_1 k_{1n} + \lambda_2 k_{2n} = 0 \\ \frac{\partial L}{\partial \lambda_1}(\mathbf{x}, \lambda_1, \lambda_2) &= \mathbf{k}_1 \cdot \mathbf{x} - c_1 = 0 \\ \frac{\partial L}{\partial \lambda_2}(\mathbf{x}, \lambda_1, \lambda_2) &= \mathbf{k}_2 \cdot \mathbf{x} - c_2 = 0\end{aligned}$$

(7.27) BEISPIEL:

Wir betrachten die Funktion  $f(x_1, x_2) = x_1^2 + x_2^2$  unter den Nebenbedingungen  $x_1 + x_2 = 1$  und  $3x_1 - 2x_2 = 5$ . Wir stellen uns die Aufgabe:

Minimiere  $x_1^2 + x_2^2$  unter der Nebenbedingung  $x_1 + x_2 = 1$ ,  $3x_1 - 2x_2 = 5$ !

Wir lösen die Aufgabe mit der Methode von Lagrange.

Die Lagrangefunktion lautet:

$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda_1(x_1 + x_2 - 1) + \lambda_2(3x_1 - 2x_2 - 5)$$

Die Bildung der partiellen Ableitungen ergibt:

$$\begin{aligned}\frac{\partial L}{\partial x_1} &= 2x_1 + \lambda_1 + 3\lambda_2 \\ \frac{\partial L}{\partial x_2} &= 2x_2 + \lambda_1 - 2\lambda_2 \\ \frac{\partial L}{\partial \lambda_1} &= x_1 + x_2 - 1 \\ \frac{\partial L}{\partial \lambda_2} &= 3x_1 - 2x_2 - 5\end{aligned}$$

Wir erhalten so das Gleichungssystem

$$\begin{array}{rcccc} 2x_1 & & \lambda_1 & +3\lambda_2 & = 0 \\ & 2x_2 & +\lambda_1 & -2\lambda_2 & = 0 \\ x_1 & +x_2 & & & = 1 \\ 3x_1 & -2x_2 & & & = 5 \end{array}$$

Wir bilden die Gleichungsmatrix

```
> G=cbind(c(2,0,1,3),c(0,2,1,-2),c(1,1,0,0),c(3,-2,0,0),c(0,0,1,5))
> G
      [,1] [,2] [,3] [,4] [,5]
[1,]    2    0    1    3    0
[2,]    0    2    1   -2    0
[3,]    1    1    0    0    1
[4,]    3   -2    0    0    5
```

und lösen das Gleichungssystem

```
> rref(G)$echelon.form
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0  1.40
[2,]    0    1    0    0 -0.40
[3,]    0    0    1    0 -0.64
[4,]    0    0    0    1 -0.72
```

Also lautet die Lösung  $\mathbf{x} = (1.4, -0.4)^t$ . Dieser kritische Punkt ist ein Minimum, denn die Funktion  $f$  ist konvex.

### 7.3.3 Das Basislemma

Wir haben bei der Charakterisierung von kritischen Punkten eines Optimierungsproblems in Satz 7.21 und 7.25 zur Begründung einen Sachverhalt aus der linearen Algebra verwendet, den wir in diesem Abschnitt genauer erklären wollen.

Zunächst formulieren wir diesen Sachverhalt in seiner einfachsten Form.

(7.28) LEMMA:

*Es seien  $\mathbf{a}$  und  $\mathbf{b}$  zwei Vektoren in  $\mathbb{R}^n$ . Dann sind die folgenden Aussagen äquivalent:*

1. *Die beiden Vektoren sind proportional, dh. es gibt eine Zahl  $\lambda$ , sodass  $\mathbf{b} = \lambda\mathbf{a}$ .*
2. *Für jeden Vektor  $\mathbf{x} \in \mathbb{R}^n$  gilt:  $\mathbf{a} \cdot \mathbf{x} = 0 \implies \mathbf{b} \cdot \mathbf{x} = 0$*

Die Implikation (1)  $\implies$  (2) ist offensichtlich und nicht der Rede wert. Die interessante Aussage ist die Umkehrung (2)  $\implies$  (1). Diese Umkehrung haben wir beim Beweis von Satz 7.21 verwendet.

Wir beweisen das Lemma auf zwei verschiedenen Wegen. Der erste Weg ist sehr einfach und direkt. Er kann aber nicht auf allgemeinere Fälle übertragen werden. Der zweite Weg ist ein wenig raffinierter, er verwendet unsere Kenntnisse über das Gaußsche Eliminationsverfahren und ist verallgemeinerbar.

BEGRÜNDUNG: (Erster Beweis)

Wir setzen (2) voraus und wollen (1) nachweisen.

Fall 1: Alle Komponenten von  $\mathbf{a}$  seien gleich 0. Wir wählen  $\mathbf{x} = (1, 0, \dots, 0)^t$ . Dann ist  $\mathbf{a} \cdot \mathbf{x} = a_1 = 0$  und daher  $b_1 = \mathbf{b} \cdot \mathbf{x} = 0$ . Das kann man mit allen Komponenten so machen, und daher ist  $\mathbf{b} = \mathbf{0}$ . Insbesondere ist dann  $\mathbf{b}$  zu  $\mathbf{a}$  proportional.

Fall 2: Es gebe zumindest eine Komponente von  $\mathbf{a}$ , die von Null verschieden ist, z.B. sei  $a_1 \neq 0$ . Wir wählen  $\mathbf{x} = (a_2/a_1, -1, 0, \dots, 0)$ . Dann ist

$$\mathbf{a} \cdot \mathbf{x} = \frac{a_2}{a_1}a_1 - a_2 = 0$$

und daher ist auch

$$\frac{a_2}{a_1}b_1 - b_2 = 0 = \mathbf{b} \cdot \mathbf{x} = 0$$

Also ist

$$b_2 = \frac{b_1}{a_1}a_2.$$

Hier kann man den Index 2 durch jeden anderen Index  $i \neq 1$  ersetzen und erhält so

$$\mathbf{b} = \frac{b_1}{a_1} \mathbf{a}.$$

□

BEGRÜNDUNG: (Zweiter Beweis)

Die Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  sind dann proportional, wenn das Gleichungssystem

$$\begin{aligned} a_1 y &= b_1 \\ a_2 y &= b_2 \\ \vdots &= \\ a_n y &= b_n \end{aligned}$$

lösbar ist. Wenn wir wie im ersten Beweis o.B.d.A. unterstellen, dass  $a_1 \neq 0$  ist, dann führt das Eliminationsverfahren auf eine Zeilenskelettform der Gestalt

$$\begin{aligned} a_1 y &= b_1 \\ 0 &= * \\ \vdots &= \\ 0 &= * \end{aligned}$$

Die Lösbarkeit ist dann gegeben, wenn die \*'s auf der rechten Seite alle gleich Null sind. Dies folgt aber aus der Bedingung (2): Jede Zeilenoperation, die die linke Seite auf Null setzt, ist eine Linearkombination der Komponenten von  $\mathbf{a}$ , und daher muss auch die entsprechende Linearkombination der Komponenten von  $\mathbf{b}$  auf der rechten Seite verschwinden. □

Nun formulieren wir den gleichen Sachverhalt für zwei Nebenbedingungen.

(7.29) LEMMA:

*Es seien  $\mathbf{a}_1, \mathbf{a}_2$  und  $\mathbf{b}$  drei Vektoren in  $\mathbb{R}^n$ . Dann sind die folgenden Aussagen äquivalent:*

1. *Es gibt Zahlen  $\lambda_1$  und  $\lambda_2$ , sodass  $\mathbf{b} = \lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2$ .*
2. *Für jeden Vektor  $\mathbf{x} \in \mathbb{R}^n$  gilt:  $\mathbf{a}_1 \cdot \mathbf{x} = 0, \mathbf{a}_2 \cdot \mathbf{x} = 0 \implies \mathbf{b} \cdot \mathbf{x} = 0$*

BEGRÜNDUNG:

Die Aussage 1. gilt dann, wenn das Gleichungssystem

$$\begin{aligned} a_{11}y_1 + a_{21}y_2 &= b_1 \\ a_{12}y_1 + a_{22}y_2 &= b_2 \\ \vdots &= \\ a_{1n}y_1 + a_{2n}y_2 &= b_n \end{aligned}$$

lösbar ist. Die Lösbarkeit ist dann gegeben, weil Bedingung (2) erfüllt ist: Jede Zeilenoperation, die die linke Seite auf Null setzt, ist eine Linearkombination der Komponenten von  $\mathbf{a}_1$  und der Komponenten von  $\mathbf{a}_2$ , und daher muss auch die entsprechende Linearkombination der Komponenten von  $\mathbf{b}$  auf der rechten Seite verschwinden. □

## 7.3.4 Aufgaben

(7.30) TESTFRAGEN

1. Durch welche Eigenschaft ist ein kritischer Punkt einer Optimierungsaufgabe mit einer linearen Nebenbedingung definiert ?
2. Durch welche Eigenschaft ist ein kritischer Punkt einer Optimierungsaufgabe mit zwei linearen Nebenbedingungen definiert ?
3. Mit welcher Methode findet man kritische Punkte von Optimierungsaufgaben mit linearen Nebenbedingungen ?
4. Erläutern Sie die Methode von Lagrange. Wozu dient sie, und wie läuft sie ab ?
5. Wie interpretiert man kritische Punkte von Optimierungsaufgaben mit linearen Nebenbedingungen ?

## 7.4 Portfolio-Optimierung

### 7.4.1 Grundlagen

#### Der Datensatz

Für die Bearbeitung von Aufgaben zur Portfolio-Optimierung stellen wir einen Datensatz zur Verfügung, der die Wochenrenditen von Aktien enthält.

Der Datensatz ist im Speicherformat von R abgespeichert worden (mit der Funktion **save**). Er kann daher nicht mit einem Texteditor gelesen werden, sondern er muss mit der Funktion **load** nach R geladen werden. Zu diesem Zweck muss sich das File **stocks.sav** im Arbeitsverzeichnis befinden:

```
> load("stocks.sav")
> attach(stocks)
> names(stocks)
 [1] "BA"    "DIS"   "GE"    "GM"    "HPQ"   "IBM"   "INTC"  "JPM"
 [9] "MSFT" "PG"    "XOM"
```

Die Daten stammen von <http://finance.yahoo.com/>. Es handelt sich um Aktien aus dem Dow-Jones Index. Im einzelnen bedeuten die Symbole:

BA	Boeing
DIS	Disney
GE	General Electric
GM	General Motors
HPQ	Hewlett Packard
IBM	IBM
INTC	Intel
JPM	JP Morgan Chase
MSFT	Microsoft
PG	Procter and Gamble
XOM	Exxon Mobile

Die Daten umfassen den Zeitraum 1.1.1990 bis 31.12.2003. Das sind 729 Wochen. Die Rohdaten in <http://finance.yahoo.com/> sind Schlusspreise der Aktien am Ende jeder Woche. Sie werden dort bereits in bereinigter Form zur Verfügung gestellt (dh. Splittings sind

berücksichtigt). Die Daten in **stocks.sav** enthalten die Renditen nach der Formel

$$\text{Rendite} = \frac{\text{Neuer Preis} - \text{Alter Preis}}{\text{Alter Preis}}$$

## Erste Schritte

---

Den Verlauf der Zeitreihe der Rendite veranschaulicht man durch ein einfaches Liniendiagramm mit der Funktion **plot**. Damit allerdings die Beschriftung des Diagramms sinnvoll ist, wandeln wir die Datenliste vorher in eine Zeitreihe um:

```
> Ba=ts(BA,start=1990,deltat=1/52)
> Ba[1:10]
 [1] -0.030545455  0.015003751 -0.049519586  0.075427683
 [5]  0.002169197  0.023088023 -0.045839210  0.081300813
 [9]  0.044429255  0.029450262
> plot(Ba)
```

Wenn wir nur einen Ausschnitt sehen wollen, dann verwenden wir die Funktion **window**:

```
> plot(window(Ba, start=c(1995,1),end=c(1995,10)))
```

Uns interessiert der Mittelwert und die Varianz (Volatilität) der Renditen. Auf diesen beiden Größen beruht die klassische Portfolio-Analyse. Natürlich schwanken diese Größen im Laufe der Zeit. Für unsere Zwecke unterstellen wir aber der Einfachheit halber, dass wir die langfristigen Durchschnittswerte die wesentliche Information enthalten:

```
> m=apply(stocks,2,mean)
> cbind(m)
```

```

              m
BA  0.002551165
DIS  0.002567259
GE   0.006254305
GM   0.002216734
HPQ  0.005344378
IBM  0.004088728
INTC 0.006759386
JPM  0.003203913
MSFT 0.006366632
PG   0.003952897
XOM  0.004259628
```

Ebenso verschaffen wir uns die Standardabweichungen:

```
> s=apply(stocks,2,sd)
> cbind(s)
```

```

              s
BA  0.04449349
DIS  0.04226970
GE   0.03984319
GM   0.04310507
HPQ  0.06062377
IBM  0.04419340
INTC 0.05894746
JPM  0.05352986
```

```
MSFT 0.04832826
PG    0.03702869
XOM   0.02905900
```

Wir sehen daraus, dass die verschiedenen Aktien durchaus völlig unterschiedliche Eigenschaften haben. Das wird anschaulich sehr deutlich, wenn wir Standardabweichungen und Mittelwerte in einem Diagramm auftragen.

```
> symbols=dimnames(stocks)[[2]]
> symbols
 [1] "BA"    "DIS"   "GE"    "GM"    "HPQ"   "IBM"   "INTC"  "JPM"
 [9] "MSFT" "PG"    "XOM"
```

```
> plot(s,m)
> text(s,m,symbols,pos=1,xpd=NA)
```

Offenbar hat XOM das geringste Risiko (kleinste Varianz), aber eine mäßige Rendite. GE, MSFT und INTC haben höhere Renditen, aber mit beträchtlich größerem Risiko.

## Portfolios

---

Wenn man ein Kapital in vorgegebener Höhe auf mehrere Aktien aufteilt, dann spricht man von einem Portfolio. Eine Grundaufgabe besteht darin, die Rendite eines Portfolios auszurechnen, wenn man nur die Renditen der Bestandteile kennt, aber nicht deren Preise.

Es seien A, B und C drei Aktien, und ihre Preise zu den Zeitpunkten  $t = 0$  und  $t = 1$  seien bekannt. Dann betragen die Renditen

$$R_A = \frac{A_1 - A_0}{A_0}, \quad R_B = \frac{B_1 - B_0}{B_0}, \quad R_C = \frac{C_1 - C_0}{C_0}.$$

Wir bilden nun aus einem Kapital in der Höhe  $K$  ein Portfolio, indem wir den Betrag  $K_1$  in die Aktie A investieren, den Betrag  $K_2$  in die Aktie B, und den Betrag  $K_3$  in die Aktie C. Klarerweise gilt  $K_1 + K_2 + K_3 = K$ . Es ist zu beachten, dass die Kapitalanteile auch negativ sein können. Solche Investitionen würden Leerverkäufen entsprechen.

Das Portfolio nennen wir P. Sein Wert zum Zeitpunkt  $t = 0$  beträgt natürlich  $P_0 = K$ , denn soviel Geld haben wir investiert. Nun rechnen wir den Wert des Portfolios zum Zeitpunkt  $t = 1$  aus. Zu diesem Zweck müssen wir die von uns gekauften Stückzahlen mit den Preisen zum Zeitpunkt  $t = 1$  multiplizieren:

$$P_1 = \frac{K_1}{A_0} A_1 + \frac{K_2}{B_0} B_1 + \frac{K_3}{C_0} C_1.$$

Die Rendite unseres Portfolios beträgt daher

$$\begin{aligned} R_P &= \frac{P_1 - P_0}{P_0} = \frac{P_1}{P_0} - 1 = \frac{P_1}{K} - 1 \\ &= \frac{K_1}{K} \frac{A_1}{A_0} + \frac{K_2}{K} \frac{B_1}{B_0} + \frac{K_3}{K} \frac{C_1}{C_0} - 1 \\ &= \frac{K_1}{K} \frac{A_1 - A_0}{A_0} + \frac{K_2}{K} \frac{B_1 - B_0}{B_0} + \frac{K_3}{K} \frac{C_1 - C_0}{C_0} \\ &= w_1 R_A + w_2 R_B + w_3 R_C \end{aligned}$$

Mit  $w_1, w_2, w_3$  bezeichnen wir die **Wertanteile** (und **nicht** die Stückzahlen) der einzelnen Komponenten. Wir fassen zusammen:

**Die Rendite eines Aktienportfolios berechnet man als gewichtetes Mittel der Renditen der Komponenten, wobei die Gewichte die eingesetzten Wertanteile sind.**

Nun können wir darangehen, Mittelwerte und Varianzen der Renditen von Portfolios zu berechnen. Die dazu nötigen Grundlagen kennen wir aus der LV über Statistik.

(7.31) SATZ:

Es seien  $R_1, R_2, \dots, R_k$  Datenlisten (Aktienrenditen) mit dem Mittelwertvektor  $\mathbf{m} = (m_1, m_2, \dots, m_k)^t$  und der Kovarianzmatrix  $\mathbf{C} = (s_{ij})_{1 \leq i, j \leq k}$ .  
Es sei  $P = w_1 R_1 + w_2 R_2 + \dots + w_k R_k$  eine Linearkombination (Portfolio) mit den Gewichtsvektor  $\mathbf{w} = (w_1, w_2, \dots, w_k)^t$ . Dann gilt

$$m_P = \sum_{i=1}^k w_i m_i = \mathbf{m}^t \mathbf{w}$$

$$s_P^2 = \sum_{i,j=1}^k s_{ij} w_i w_j = \mathbf{w}^t \mathbf{C} \mathbf{w}$$

Wir benötigen also den Mittelwertvektor und die Kovarianzmatrix aller Aktien, die wir in ein Portfolio nehmen wollen. Es ist

```
> m=apply(stocks, 2, mean)
> C=cov(stocks)
```

Wenn wir uns für ein Portfolio entscheiden, z.B. für

```
> w=runif(11)
> w=w/sum(w)
> w
 [1] 0.197844018 0.240518182 0.062134683 0.001792784 0.031165500
 [6] 0.015990053 0.051565628 0.025152590 0.075548602 0.057339765
[11] 0.240948195
```

Dann berechnen wir Mittelwert und Standardabweichung der Renditen des Portfolios durch

```
> sum(m*w)
[1] 0.003909864
> sqrt(rbind(w)%*%C*%cbind(w))
```

```
          w
w 0.0255127
```

Natürlich ist es unbefriedigend, unsere Portfolios zufällig zusammenzustellen. Wir wollen vielmehr einen systematischen Gesichtspunkt für die Konstruktion eines Portfolios verwenden. Das werden wir im nächsten Abschnitt machen.

Zum Abschluss dieses Abschnitts berechnen wir noch eine grosse Zahl von zufälligen Portfolios, um zu sehen, welche Mittelwerte und Standardabweichungen mit Portfolios überhaupt erzielbar sind.

Die Funktion **rportfolio** aus der Library **LinAlg.r** erzeugt zufällige Portfolios und berechnet ihre Mittelwerte und Varianzen. Anschliessend zeichnen wir die Mittelwerte und Standardabweichungen in ein Diagramm ein:

```
> p=rportfolio(10000,m,C)
> plot(s,m,xlim=c(0.02,0.065),pch=19)
```

```
> points(sqrt(p$s), p$m, pch=".")
> text(s, m, symbols, pos=1, xpd=NA, font=2)
```

Wie sehen, dass es Portfolios gibt, die wesentlich geringere Varianzen aufweisen als Einzelaktien mit der gleichen mittleren Rendite. Das ist die Folge der **Diversifikation**.

## 7.4.2 Optimale Portfolios

Das Ziel eines Investors besteht darin, mit möglichst geringem Risiko eine möglichst hohe Rendite zu erzielen. Daher gibt es einen Zielkonflikt (trade-off) zwischen Minimierung des Risikos und Maximierung der Rendite.

### Vorbemerkung

Bei der folgenden Portfolio-Optimierung wird immer wieder die quadratische Funktion  $f(\mathbf{w}) = \mathbf{w}^t \mathbf{C} \mathbf{w}$  auftreten. Es ist dabei von wesentlicher Bedeutung zu klären, ob diese Funktion konvex ist.

Im vorliegenden Fall ist die Frage deshalb sehr einfach zu beantworten, weil die Matrix  $\mathbf{C}$  eine Kovarianzmatrix ist.

(7.32) SATZ:

*Ist  $\mathbf{C}$  eine Kovarianzmatrix, dann ist die quadratische Funktion  $f(\mathbf{w}) = \mathbf{w}^t \mathbf{C} \mathbf{w}$  konvex.*

BEGRÜNDUNG: Die Funktion  $f(\mathbf{w})$  gibt immer die Varianz einer Linearkombination von Datenlisten an. Da eine Varianz stets nichtnegativ ist, hat die Funktion  $f(\mathbf{w})$  nur nichtnegative Werte. Daher kann sie nur konvex sein, weil in jedem anderen Fall (durch die Existenz negativer Eigenwerte) sicher auch negative Funktionswerte auftreten würden.  $\square$

### Minimierung der Varianz

Die einfachste Aufgabe besteht darin, das Portfolio mit der kleinsten Varianz zu ermitteln:

**Minimiere  $\mathbf{w}^t \mathbf{C} \mathbf{w}$  unter der Bedingung  $\mathbf{e}^t \mathbf{w} = 1$ .**

Es handelt sich also um ein quadratisches Optimierungsproblem mit einer linearen Nebenbedingung. Da die zugrundeliegende quadratische Funktion konvex ist, erhält man Minima durch Berechnung der kritischen Punkte.

Wir stellen die Lagrangefunktion auf:

$$L(\mathbf{w}, \lambda) = \mathbf{w}^t \mathbf{C} \mathbf{w} + \lambda(\mathbf{e}^t \mathbf{w} - 1)$$

Wir bilden partielle Ableitungen nach  $\mathbf{w}$  und  $\lambda$ :

$$\begin{aligned}\nabla_{\mathbf{w}} L &= 2\mathbf{C}\mathbf{w} + \lambda\mathbf{e} \\ \nabla_{\lambda} L &= \mathbf{e} \cdot \mathbf{w} - 1\end{aligned}$$

Um kritische Punkte zu finden, müssen wir also das Gleichungssystem

$$\begin{aligned}2\mathbf{C}\mathbf{w} + \lambda\mathbf{e} &= 0 \\ \mathbf{e}^t \mathbf{w} &= 1\end{aligned}$$

mit den  $k+1$  Unbekannten  $w_1, w_2, \dots, w_k, \lambda$  lösen. Man nennt dieses Gleichungssystem das System der **Normalgleichungen**. Jede Lösung dieses Gleichungssystems ist ein kritischer Punkt. Das Gleichungssystem hat die Gleichungsmatrix

$$\left( \begin{array}{cccc|c} 2s_{11} & 2s_{12} & \dots & 2s_{1k} & 1 & 0 \\ 2s_{12} & 2s_{22} & \dots & 2s_{2k} & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 2s_{1k} & 2s_{2k} & \dots & 2s_{kk} & 1 & 0 \\ 1 & 1 & \dots & 1 & 0 & 1 \end{array} \right)$$

Da  $f$  konvex ist, können wir den gewonnenen kritischen Punkt sofort als Minimum interpretieren.

(7.33) BEISPIEL:

Wir illustrieren die Portfolio-Optimierung am Beispiel von Portfolios, die aus drei Aktien DIS, IBM und PG bestehen.

Zunächst berechnen wir die Mittelwerte

```
> xx=cbind(DIS,IBM,PG)
> m=apply(xx,2,mean)
> cbind(m)
```

```
          m
DIS 0.002567259
IBM 0.004088728
PG  0.003952897
```

und die Kovarianzmatrix

```
> C=cov(xx)
> C
          DIS          IBM          PG
DIS 0.0017867272 0.0005412995 0.0004004644
IBM 0.0005412995 0.0019530568 0.0001657983
PG  0.0004004644 0.0001657983 0.0013711238
```

Anschließend bauen wir die Gleichungsmatrix der Normalgleichungen auf:

```
> e=c(1,1,1)
> G=cbind(2*C,e,rep(0,3))
> G=rbind(G,c(e,0,1))
> G
          DIS          IBM          PG e
DIS 0.0035734544 0.0010825991 0.0008009287 1 0
IBM 0.0010825991 0.0039061135 0.0003315965 1 0
PG  0.0008009287 0.0003315965 0.0027422476 1 0
      1.0000000000 1.0000000000 1.0000000000 0 1
```

Das Eliminationsverfahren

```
> r=rref(G)$echelon.form
> r
          DIS IBM PG e
DIS      1   0  0  0  0.24464490
```

```

IBM    0    1    0    0    0.29272166
PG     0    0    1    0    0.46263343
      0    0    0    1   -0.00156166

```

liefert uns die Lösung

```

> w=r[1:3,5]
> w
      DIS      IBM      PG
0.2446449 0.2927217 0.4626334

```

Das sind die Gewichte des Portfolios mit der geringsten Varianz. Mittelwert und Varianz dieses Portfolios betragen

```

> sum(m*w)
[1] 0.003653668
> rbind(w)%*%C%*%cbind(w)

      w
w 0.000780832

```

## Minimierung der Varianz bei vorgegebener Rendite

Unsere nächste Aufgabe besteht darin, ein Portfolio mit der kleinsten Varianz zu ermitteln, das darüber hinaus eine vorgegebene Rendite  $\mu$  besitzt:

**Minimiere  $\mathbf{w}^t \mathbf{C} \mathbf{w}$  unter den Bedingungen  $\mathbf{e}^t \mathbf{w} = 1$  und  $\mathbf{m}^t \mathbf{w} = \mu$ .**

Es handelt sich um ein quadratisches Optimierungsproblem mit zwei linearen Nebenbedingungen. Da die zugrundeliegende quadratische Funktion wieder konvex ist, erhält man Minima durch Berechnung der kritischen Punkte.

Wir stellen die Lagrangefunktion auf:

$$L(\mathbf{w}, \lambda_1, \lambda_2) = \mathbf{w}^t \mathbf{C} \mathbf{w} + \lambda_1 (\mathbf{e}^t \mathbf{w} - 1) + \lambda_2 (\mathbf{m}^t \mathbf{w} - \mu)$$

Wir bilden partielle Ableitungen nach  $\mathbf{w}$ ,  $\lambda_1$  und  $\lambda_2$ :

$$\begin{aligned} \nabla_{\mathbf{w}} L &= 2\mathbf{C}\mathbf{w} + \lambda_1 \mathbf{e} + \lambda_2 \mathbf{m} \\ \nabla_{\lambda_1} L &= \mathbf{e} \cdot \mathbf{w} - 1 \\ \nabla_{\lambda_2} L &= \mathbf{m} \cdot \mathbf{w} - \mu \end{aligned}$$

Um kritische Punkte zu finden, müssen wir also das Gleichungssystem

$$\begin{aligned} 2\mathbf{C}\mathbf{w} + \lambda_1 \mathbf{e} + \lambda_2 \mathbf{m} &= 0 \\ \mathbf{e}^t \mathbf{w} &= 1 \\ \mathbf{m}^t \mathbf{w} &= \mu \end{aligned}$$

mit den  $k + 2$  Unbekannten  $w_1, w_2, \dots, w_k, \lambda_1, \lambda_2$  lösen. Jede Lösung dieses Gleichungssystems ist ein kritischer Punkt. Das Gleichungssystem hat die Gleichungsmatrix

$$\left( \begin{array}{cccccc|c} 2s_{11} & 2s_{12} & \dots & 2s_{1k} & 1 & m_1 & 0 \\ 2s_{12} & 2s_{22} & \dots & 2s_{1k} & 1 & m_2 & 0 \\ \dots & & & & & & \\ 2s_{1k} & 2s_{2k} & \dots & 2s_{kk} & 1 & m_k & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 & 1 \\ m_1 & m_2 & \dots & m_k & 0 & 0 & \mu \end{array} \right)$$

Da  $f$  konvex ist, können wir den gewonnenen kritischen Punkt sofort als Minimum interpretieren.

(7.34) BEISPIEL:

Wir wollen nun ein Portfolio bestimmen, bei dem die durchschnittliche Rendite  $\mu = 0.0035$  beträgt, und das unter dieser Nebenbedingung minimale Varianz besitzt.

Die Gleichungsmatrix der Normalgleichungen lautet

```
> G=cbind(2*C,e,m,rep(0,3))
> G=rbind(G,c(e,0,0,1),c(m,0,0,0.0035))
> G
```

	DIS	IBM	PG	e	m
DIS	0.0035734544	0.0010825991	0.0008009287	1	0.002567259
IBM	0.0010825991	0.0039061135	0.0003315965	1	0.004088728
PG	0.0008009287	0.0003315965	0.0027422476	1	0.003952897
e	1.0000000000	1.0000000000	1.0000000000	0	0.0000000000
m	0.0025672588	0.0040887284	0.0039528970	0	0.0000000000

und das Eliminationsverfahren liefert wieder die Lösung:

```
> r=rref(G)$echelon.form
> r
```

	DIS	IBM	PG	e	m
DIS	1	0	0	0	0.35031684
IBM	0	1	0	0	0.23938068
PG	0	0	1	0	0.41030248
e	0	0	0	1	-0.00249644
m	0	0	0	0	0.25584601

```
> w=r[1:3,6]
> w
```

	DIS	IBM	PG
w	0.3503168	0.2393807	0.4103025

Dieses Portfolio hat Mittelwert und Varianz

```
> sum(m*w)
[1] 0.0035
> rbind(w)%*%C%*%cbind(w)
```

	w
w	0.0008004897

## Maximierung der Rendite

---

Die auf den ersten Blick attraktivste Optimierungsaufgabe ist natürlich die Maximierung der Rendite. Aber wie uns schon der Blick auf unsere Diagramme gelehrt hat, ist eine Renditesteigerung in der Regel mit einer Risikoerhöhung verbunden.

Ein einfaches Nutzenmodell unterstellt, dass der Nutzen eines Portfolios mit dem Mittelwert  $\mu$  und der Varianz  $\sigma^2$  für einen Investor durch den Ausdruck

$$U = \mu - \gamma\sigma^2$$

beschrieben werden kann, wobei  $\gamma > 0$  die sogenannte **Risikoaversion** des Investors ausdrückt. Der Nutzen hat also die Form

$$U(\mathbf{w}) = \mathbf{m}^t \mathbf{w} - \gamma \mathbf{w}^t \mathbf{C} \mathbf{w}.$$

Das ist eine konkave quadratische Funktion. Daher ist es möglich, Portfolios zu berechnen, die den so definierten Nutzen maximieren.

Die Aufgabe besteht also darin, folgendes Maximumproblem zu lösen:

$$\text{Maximiere } \mathbf{m}^t \mathbf{w} - \gamma \mathbf{w}^t \mathbf{C} \mathbf{w} \text{ unter der Bedingung } \mathbf{e}^t \mathbf{w} = 1.$$

Dabei ist  $\gamma > 0$  eine Konstante, deren Wert gegeben ist. Es handelt sich also um ein quadratisches Optimierungsproblem mit einer linearen Nebenbedingung. Da die zugrundeliegende quadratische Funktion konkav ist, erhält man Maxima durch Berechnung der kritischen Punkte.

Wir stellen die Lagrangefunktion auf:

$$L(\mathbf{w}, \lambda) = \mathbf{m}^t \mathbf{w} - \gamma \mathbf{w}^t \mathbf{C} \mathbf{w} + \lambda(\mathbf{e}^t \mathbf{w} - 1)$$

Wir bilden partielle Ableitungen nach  $\mathbf{w}$  und  $\lambda$ :

$$\begin{aligned} \nabla_{\mathbf{w}} L &= -2\gamma \mathbf{C} \mathbf{w} + \mathbf{m} + \lambda \mathbf{e} \\ \nabla_{\lambda} L &= \mathbf{e} \cdot \mathbf{w} - 1 \end{aligned}$$

Um kritische Punkte zu finden, müssen wir also das Gleichungssystem

$$\begin{aligned} -2\gamma \mathbf{C} \mathbf{w} + \lambda \mathbf{e} &= -\mathbf{m} \\ \mathbf{e}^t \mathbf{w} &= 1 \end{aligned}$$

mit den  $k+1$  Unbekannten  $w_1, w_2, \dots, w_k, \lambda$  lösen. Jede Lösung dieses Gleichungssystems ist ein kritischer Punkt. Das Gleichungssystem hat die Gleichungsmatrix

$$\left( \begin{array}{cccc|c} -2\gamma s_{11} & -2\gamma s_{12} & \dots & -2\gamma s_{1k} & 1 & -m_1 \\ -2\gamma s_{12} & -2\gamma s_{22} & \dots & -2\gamma s_{1k} & 1 & -m_2 \\ \dots & & & & & \\ -2\gamma s_{1k} & -2\gamma s_{2k} & \dots & -2\gamma s_{kk} & 1 & -m_k \\ 1 & 1 & \dots & 1 & 0 & 1 \end{array} \right)$$

Da  $f$  konkav ist, können wir den gewonnenen kritischen Punkt sofort als Maximum interpretieren.

(7.35) BEISPIEL:

Wir maximieren den Nutzen für einen Investor mit der Risikoaversion  $\gamma = 2$ .

Die Gleichungsmatrix der Normalgleichungen lautet

```

> gamma=2
> G=cbind(-2*gamma*C,e,-m)
> G=rbind(G,c(e,0,1))
> G
              DIS              IBM              PG e
DIS -0.007146909 -0.002165198 -0.001601857 1 -0.002567259
IBM -0.002165198 -0.007812227 -0.000663193 1 -0.004088728
PG  -0.001601857 -0.000663193 -0.005484495 1 -0.003952897
      1.000000000  1.000000000  1.000000000 0  1.000000000

```

und das Eliminationsverfahren ergibt

```

> r=rref(G)$echelon.form
> r
      DIS IBM PG e
DIS   1   0  0  0  0.03813019
IBM   0   1  0  0  0.39696597
PG    0   0  1  0  0.56490384
      0   0  0  1 -0.00053034
> w=r[1:3,5]
> w
      DIS              IBM              PG
0.03813019 0.39696597 0.56490384

```

Die Kenngrößen der Lösung betragen

```

> sum(m*w)
[1] 0.003953983
> rbind(w)%*%C%*%cbind(w)
              w
w 0.0008559106

```

## 7.4.3 Aufgaben

### (7.36) TESTFRAGEN

1. Begründen Sie, warum Kovarianzmatrizen immer zu konvexen quadratischen Funktionen führen.
2. Wie lauten die Optimierungsprobleme in Zusammenhang mit der Portfolio-Optimierung?