

Anhang A

Rechnen mit Matrizen

Die Darstellung einiger der in diesem Text behandelten Verfahren wird durch die Verwendung der Matrizenrechnung wesentlich erleichtert. In diesem Anhang erklären wir die erforderlichen Grundbegriffe.

A.1 Grundbegriffe

1. Unter einer *Matrix* verstehen wir eine rechteckig angeordnete Menge von Zahlen, wobei es sich im allgemeinen um beliebige reelle Zahlen handeln kann. Wenn es n Zeilen und m Spalten gibt, spricht man von einer (n, m) -Matrix oder von einer Matrix der *Ordnung* (n, m) . Zum Beispiel ist

$$\mathbf{A} := \begin{pmatrix} 1 & 3 & 9 \\ 4 & 1 & 7 \\ 2 & 7 & 8 \\ 2 & 1 & 5 \end{pmatrix}$$

eine $(4, 3)$ -Matrix. Die Schreibweise $\mathbf{A} := \dots$ sagt, daß man dieser Matrix den Namen \mathbf{A} geben möchte. Beliebige andere Namen sind möglich; in diesem Text verwenden wir (meistens) fettgedruckte Buchstaben.¹

2. Um allgemein auf eine (n, m) -Matrix Bezug nehmen zu können, verwenden wir die Schreibweise

$$\mathbf{A} = (a_{ij})$$

wobei sich der Index i auf die Zeilen, der Index j auf die Spalten der Matrix bezieht. In ausführlicher Schreibweise hat man also

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

Das Matrixelement a_{ij} steht in der i -ten Zeile und j -ten Spalte. Wenn \mathbf{A} eine Matrix ist, verwenden wir gelegentlich auch die Schreibweise $(\mathbf{A})_{ij}$, um auf das Matrixelement Bezug zu nehmen, das sich in der i -ten Zeile und j -ten Spalte von \mathbf{A} befindet.

¹Eine Ausnahme ist das Symbol \mathbf{R} , mit dem stets auf die Menge der reellen Zahlen verwiesen wird.

3. *Vektoren* sind Matrizen, die nur aus einer Zeile oder einer Spalte bestehen. $(1, m)$ -Matrizen werden als *Zeilenvektoren*, $(n, 1)$ -Matrizen werden als *Spaltenvektoren* bezeichnet. Im Kontext der Matrizenrechnung werden $(1, 1)$ -Matrizen als *Skalare* bezeichnet und mit dem entsprechenden Matrixelement identifiziert. Zur Bezeichnung von Vektoren verwenden wir (meistens) fettgedruckte Kleinbuchstaben. Um uns auf die Spaltenvektoren einer (n, m) -Matrix \mathbf{A} zu beziehen, schreiben wir oft

$$\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_m)$$

Gemeint ist, daß \mathbf{a}_1 die erste Spalte von \mathbf{A} ist, \mathbf{a}_2 die zweite Spalte usw.

4. Es seien \mathbf{A} und \mathbf{B} zwei Matrizen der gleichen Ordnung. Wir schreiben

$$\mathbf{A} = \mathbf{B} \quad \text{wenn für alle } i, j \text{ gilt: } a_{ij} = b_{ij}$$

Analog werden die Relationen $<$ und \leq sowie $>$ und \geq definiert, z.B.:

$$\mathbf{A} \leq \mathbf{B} \quad \text{wenn für alle } i, j \text{ gilt: } a_{ij} \leq b_{ij}$$

Man beachte, daß diese Relationen nur eine partielle Ordnung der Matrizen liefern.

5. Um mit Matrizen rechnen zu können, werden einige Operationen vereinbart. Die einfachste Operation ist die Multiplikation einer Matrix mit einem Skalar. Sei $\mathbf{A} = (a_{ij})$ eine Matrix und c ein Skalar; die Definition lautet: $c\mathbf{A} := (ca_{ij})$. Es wird also jedes Element von \mathbf{A} mit der Zahl c multipliziert. Wenn $c = 0$ ist, entsteht eine *Nullmatrix*, bei der alle Elemente gleich Null sind. Wir werden für Nullmatrizen das Symbol $\mathbf{0}$ verwenden und annehmen, daß sich ihre Ordnung aus dem Kontext ergibt.

6. Matrizen, die die gleiche Ordnung haben, können addiert und subtrahiert werden. Seien \mathbf{A} und \mathbf{B} zwei (n, m) -Matrizen, dann ist

$$\mathbf{A} + \mathbf{B} := (a_{ij} + b_{ij}) \quad \text{und} \quad \mathbf{A} - \mathbf{B} := (a_{ij} - b_{ij})$$

Man wendet also die Operationen einzeln auf alle sich entsprechenden Matrixelemente an.

7. Eine Matrix \mathbf{A} kann mit einer Matrix \mathbf{B} multipliziert werden, wenn die Anzahl der Spalten von \mathbf{A} und die Anzahl der Zeilen von \mathbf{B} gleich sind. Sei also $\mathbf{A} = (a_{ij})$ eine (n, m) -Matrix und $\mathbf{B} = (b_{ij})$ eine (m, p) -Matrix. Dann ist das Produkt $\mathbf{C} := \mathbf{AB}$ eine (n, p) -Matrix mit den Elementen

$$c_{ij} := \sum_{k=1}^m a_{ik}b_{kj}$$

Zum Beispiel kann die zu Beginn definierte $(4, 3)$ -Matrix \mathbf{A} mit einem

(3,1)-Spaltenvektor $\mathbf{b} := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ multipliziert werden. Man erhält als

Ergebnis den (4,1)-Spaltenvektor $\mathbf{A}\mathbf{b} = \begin{pmatrix} 13 \\ 12 \\ 17 \\ 8 \end{pmatrix}$, der die Zeilensummen

der Matrix \mathbf{A} enthält. Man beachte, daß die Matrixmultiplikation nicht kommutativ ist, d.h. im allgemeinen ist $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$.

8. Unter Verwendung der bisher gegebenen Definitionen kann man sich davon überzeugen, daß folgende Regeln für das Rechnen mit Matrizen gültig sind. Dabei sind \mathbf{A} , \mathbf{B} und \mathbf{C} jeweils Matrizen passender Ordnung, und a und b sind Skalare.

$$\begin{array}{ll} \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} & a(\mathbf{B} + \mathbf{C}) = a\mathbf{B} + a\mathbf{C} \\ \mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C} & a(\mathbf{B} - \mathbf{C}) = a\mathbf{B} - a\mathbf{C} \\ \mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{A}\mathbf{B})\mathbf{C} & (a+b)\mathbf{C} = a\mathbf{C} + b\mathbf{C} \\ \mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C} & (a-b)\mathbf{C} = a\mathbf{C} - b\mathbf{C} \\ (\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{B}\mathbf{A} + \mathbf{C}\mathbf{A} & a(b\mathbf{C}) = (ab)\mathbf{C} \\ \mathbf{A}(\mathbf{B} - \mathbf{C}) = \mathbf{A}\mathbf{B} - \mathbf{A}\mathbf{C} & a(\mathbf{B}\mathbf{C}) = (a\mathbf{B})\mathbf{C} \\ (\mathbf{B} - \mathbf{C})\mathbf{A} = \mathbf{B}\mathbf{A} - \mathbf{C}\mathbf{A} & a\mathbf{B} = \mathbf{B}a \end{array}$$

9. Eine oft verwendete Operation ist die *Transposition*, die darin besteht, daß die Zeilen und Spalten einer Matrix vertauscht werden. Sei $\mathbf{A} = (a_{ij})$ eine (n, m) -Matrix. Durch Transposition entsteht eine (m, n) -Matrix

$$\mathbf{A}' := (a_{ji})$$

Wir verwenden stets ein einfaches Hochkomma, um die Transponierte einer Matrix zu bezeichnen. Es gilt folgende wichtige Regel:

$$(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$$

Die Transposition ist auch für das Rechnen mit Vektoren wichtig. Seien \mathbf{x} und \mathbf{y} zwei $(n, 1)$ -Spaltenvektoren. Dann ist $\mathbf{x}'\mathbf{y}$ ein Skalar, jedoch $\mathbf{y}\mathbf{x}'$ eine (n, n) -Matrix.

10. Wenn $\mathbf{x} = (x_1, \dots, x_n)'$ ein Spaltenvektor ist, kann man insbesondere das Produkt $\mathbf{x}'\mathbf{x} = \sum_{i=1}^n x_i^2$ bilden. Das Ergebnis ist eine stets nicht-negative Zahl, deren Wurzel

$$\|\mathbf{x}\| := \sqrt{\mathbf{x}'\mathbf{x}}$$

als *euklidische Länge* des Vektors \mathbf{x} bezeichnet wird.

11. Eine Matrix heißt *quadratisch*, wenn sie gleich viele Zeilen und Spalten hat. Ist $\mathbf{A} = (a_{ij})$ eine quadratische (n, n) -Matrix, ist ihre *Spur* (*trace*) durch

$$\text{tr}(\mathbf{A}) := \sum_{i=1}^n a_{ii}$$

definiert. Es gilt folgender Zusammenhang:

$$\text{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 = \text{tr}(\mathbf{A}\mathbf{A}')$$

Diese Gleichung gilt auch für beliebige (m, n) -Matrizen.

12. Eine quadratische Matrix \mathbf{A} heißt *symmetrisch*, wenn sie mit ihrer Transponierten identisch ist, wenn also $\mathbf{A} = \mathbf{A}'$ gilt. Man überlege sich, daß $\mathbf{A}'\mathbf{A}$ stets eine symmetrische Matrix ist.

13. Eine quadratische (n, n) -Matrix $\mathbf{A} = (a_{ij})$ wird eine *Diagonalmatrix* genannt, wenn alle Elemente außerhalb der Hauptdiagonalen Null sind. Zur Definition von Diagonalmatrizen verwenden wir oft die Schreibweise:

$$\text{diag}(a_1, \dots, a_n) := \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix}$$

Offenbar ist jede Diagonalmatrix auch symmetrisch.

14. Ein wichtiges Beispiel für eine Diagonalmatrix ist die Einheitsmatrix. Eine (n, n) -*Einheitsmatrix* ist durch

$$\mathbf{I}_n := \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

definiert; in der Hauptdiagonalen stehen Einsen und außerhalb stehen Nullen. Wenn \mathbf{A} eine beliebige (n, m) -Matrix ist, gilt offenbar: $\mathbf{A}\mathbf{I}_m = \mathbf{A}$ und $\mathbf{I}_n\mathbf{A} = \mathbf{A}$. Sei nun außerdem \mathbf{B} eine (m, p) -Matrix. Dann gilt auch: $\mathbf{A}\mathbf{I}_m\mathbf{B} = \mathbf{A}\mathbf{B}$.

A.2 Datenmatrizen

1. Die Matrizenrechnung dient uns in diesem Text hauptsächlich dafür, um auf einfache Weise mit Datenmatrizen umgehen zu können. In diesem Abschnitt definieren wir das Konzept einer Datenmatrix und erinnern kurz an einige Begriffe der Statistik. — Ausgangspunkt für die Konzeption von

Datenmatrizen sind statistische Variablen. Nehmen wir an, daß man sich auf m statistische Variablen X_1, \dots, X_m beziehen möchte und daß es Werte dieser Variablen für n Objekte (Individuen, Situationen) gibt. Dann kann man mit x_{ij} den Wert der Variablen X_j beim Objekt i bezeichnen und diese Werte in einer Datenmatrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

zusammenfassen. Die j -te Spalte dieser Matrix enthält die Werte der Variablen X_j , die i -te Zeile enthält die Werte der Variablen X_1, \dots, X_m für das Objekt i .

2. Der Mittelwert einer Variablen X_j ist durch

$$M(X_j) := \frac{1}{n} \sum_{i=1}^n x_{ij}$$

definiert. Sei jetzt $\mathbf{1}_n$ ein $(n, 1)$ -Spaltenvektor, bei dem alle Elemente gleich 1 sind. Dann kann man in folgender Weise einen Zeilenvektor bilden, der die Mittelwerte aller Variablen enthält:

$$\frac{1}{n} \mathbf{1}'_n \mathbf{X} = (M(X_1), \dots, M(X_m))$$

3. Wichtig für die Überlegungen des Haupttextes sind auch die Begriffe Varianz, Kovarianz und Korrelation. Wir erinnern deshalb kurz an die Definitionen. Die *Varianz* einer Variablen X_j ist durch

$$V(X_j) := \frac{1}{n} \sum_{i=1}^n (x_{ij} - M(X_j))^2$$

definiert. Die *Kovarianz* von zwei Variablen X_j und $X_{j'}$ ist durch

$$\text{Cov}(X_j, X_{j'}) := \frac{1}{n} \sum_{i=1}^n (x_{ij} - M(X_j))(x_{ij'} - M(X_{j'}))$$

definiert; und schließlich die *Korrelation* von X_j und $X_{j'}$ durch

$$\text{Corr}(X_j, X_{j'}) := \frac{\text{Cov}(X_j, X_{j'})}{\sqrt{V(X_j)V(X_{j'})}}$$

4. Es ist oft praktisch, die Kovarianzen einer Menge von Variablen X_1, \dots, X_m in einer *Kovarianzmatrix*

$$\text{Cov}(\mathbf{X}) := \begin{pmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_m) \\ \vdots & & \vdots \\ \text{Cov}(X_m, X_1) & \cdots & \text{Cov}(X_m, X_m) \end{pmatrix}$$

zusammenzufassen. Es ist eine symmetrische (m, m) -Matrix. Die Hauptdiagonale enthält die Varianzen. (Wir verwenden wahlweise $V(X_j)$ oder $\text{Cov}(X_j, X_j)$, um die Varianz von X_j zu bezeichnen.) Ganz analog kann man eine *Korrelationsmatrix*

$$\text{Corr}(\mathbf{X}) := \begin{pmatrix} \text{Corr}(X_1, X_1) & \cdots & \text{Corr}(X_1, X_m) \\ \vdots & & \vdots \\ \text{Corr}(X_m, X_1) & \cdots & \text{Corr}(X_m, X_m) \end{pmatrix}$$

definieren.

5. Die Darstellung von Kovarianzmatrizen wird besonders einfach, wenn man um ihre Mittelwerte *zentrierte Variablen* verwendet, also

$$Z_j := X_j - M(X_j)$$

anstelle von X_j . Für die Kovarianzen findet man dann nämlich

$$\text{Cov}(Z_j, Z_{j'}) = \frac{1}{n} \sum_{i=1}^n z_{ij} z_{ij'}$$

und es gibt die einfache Darstellung

$$\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{Z}) = \frac{1}{n} \mathbf{Z}'\mathbf{Z}$$

für die Kovarianzmatrix der Variablen Z_1, \dots, Z_m .

6. Wenn man nicht nur zentrierte, sondern *standardisierte Variablen* verwendet (vgl. Abschnitt 5.4), also

$$Z_j := \frac{X_j - M(X_j)}{\sqrt{V(X_j)}}$$

ergibt sich eine analoge Vereinfachung für die Korrelationsmatrix. Denn die standardisierten Variablen haben die Varianz 1 und die Korrelationen zwischen den standardisierten Variablen sind mit ihren Kovarianzen identisch. Also gilt bei standardisierten Variablen:

$$\text{Corr}(\mathbf{X}) = \text{Corr}(\mathbf{Z}) = \text{Cov}(\mathbf{Z}) = \frac{1}{n} \mathbf{Z}'\mathbf{Z}$$

A.3 Invertierbare Matrizen

1. Sei \mathbf{A} eine quadratische (n, n) -Matrix. \mathbf{A} heißt *invertierbar*, wenn es eine (n, n) -Matrix \mathbf{B} gibt, so daß gilt:

$$\mathbf{BA} = \mathbf{AB} = \mathbf{I}_n \tag{A.3.1}$$

Man beachte, daß nur bei quadratischen Matrizen von Invertierbarkeit gesprochen wird.

2. Nicht alle quadratischen Matrizen sind invertierbar. Die folgende Matrix

$$\mathbf{A} := \begin{pmatrix} 1 & 2 \\ 4 & 2 \end{pmatrix}$$

ist jedoch invertierbar. Denn wenn man

$$\mathbf{B} := \begin{pmatrix} -1/3 & 1/3 \\ 2/3 & -1/6 \end{pmatrix}$$

verwendet, kann man sich von der Gültigkeit der Beziehung (A.3.1) überzeugen.

3. Wenn \mathbf{A} invertierbar ist, gibt es genau eine Matrix \mathbf{B} , die die Gleichung (A.3.1) erfüllt. Denn angenommen, es gäbe noch eine zweite Matrix \mathbf{C} , so daß $\mathbf{CA} = \mathbf{AC} = \mathbf{I}_n$ ist. Dann folgt: $\mathbf{B} = \mathbf{BAC} = \mathbf{C}$. Die bei einer invertierbaren Matrix \mathbf{A} durch (A.3.1) eindeutig bestimmte Matrix \mathbf{B} wird die *zu \mathbf{A} inverse Matrix* (kurz *Inverse*) genannt und mit \mathbf{A}^{-1} bezeichnet. Also: wenn \mathbf{A} invertierbar ist, gibt es genau eine inverse Matrix \mathbf{A}^{-1} , und es gilt:

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

4. Wenn \mathbf{A} eine invertierbare Matrix ist, ist auch \mathbf{A}^{-1} invertierbar, und es gilt:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

Wenn \mathbf{A} und \mathbf{B} invertierbare (n, n) -Matrizen sind, ist auch (\mathbf{AB}) invertierbar, und es gilt:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

denn: $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n$. Wenn \mathbf{A} eine invertierbare (n, n) -Matrix ist, dann ist auch \mathbf{A}' eine invertierbare Matrix, und es gilt:

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

denn: $\mathbf{A}'(\mathbf{A}^{-1})' = (\mathbf{A}^{-1}\mathbf{A})' = \mathbf{I}_n' = \mathbf{I}_n$.

5. Bei kleinen invertierbaren Matrizen kann man ihre Inversen per Hand ausrechnen; die Verfahren sind jedoch mühselig und sollen hier nicht besprochen werden. Bei größeren Matrizen wird man stets einen Computer

verwenden, um ihre Inversen zu berechnen. Eine Ausnahme bilden Diagonalmatrizen, deren Inverse (wenn es sie gibt) direkt hingeschrieben werden können. Sei nämlich

$$\mathbf{A} := \text{diag}(a_1, \dots, a_n)$$

eine Diagonalmatrix. Dann ist \mathbf{A} genau dann invertierbar, wenn alle Koeffizienten a_1, \dots, a_n ungleich Null sind; und die Inverse ist

$$\mathbf{A}^{-1} = \text{diag}(1/a_1, \dots, 1/a_n)$$

A.4 Linearkombinationen

1. Sei $\mathbf{a}_1, \dots, \mathbf{a}_m$ eine Menge von $(n, 1)$ -Spaltenvektoren. Eine *Linearkombination* dieser Vektoren ist ein Ausdruck der Form

$$k_1\mathbf{a}_1 + \dots + k_m\mathbf{a}_m$$

wobei k_1, \dots, k_m reelle Zahlen (Skalare) sind; sie werden die *Koeffizienten* der Linearkombination genannt. Es ist klar, daß eine solche Linearkombination wiederum einen $(n, 1)$ -Spaltenvektor liefert. Eine Linearkombination wird *nicht-trivial* genannt, wenn mindestens einer der Koeffizienten ungleich Null ist.

2. Die Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_m$ heißen *linear abhängig*, wenn es eine Linearkombination

$$k_1\mathbf{a}_1 + \dots + k_m\mathbf{a}_m = \mathbf{0}$$

gibt, wobei mindestens einer der Koeffizienten ungleich Null ist. Die Vektoren heißen *linear unabhängig*, wenn sie nicht linear abhängig sind. Es folgt aus der Definition: Wenn einer der Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_m$ ein Nullvektor ist, sind die Vektoren linear abhängig. Daraus folgt umgekehrt: Wenn die Vektoren linear unabhängig sind, kann keiner der Vektoren ein Nullvektor sein. Weiterhin folgt unmittelbar aus der Definition: Wenn die Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_m$ linear unabhängig sind, ist auch jede Teilmenge dieser Vektoren linear unabhängig.

3. Wenn die Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_m$ linear abhängig sind, kann mindestens einer der Vektoren als eine Linearkombination der übrigen dargestellt werden. Denn es gibt dann eine Linearkombination

$$k_1\mathbf{a}_1 + \dots + k_m\mathbf{a}_m = \mathbf{0}$$

wobei mindestens einer der Koeffizienten ungleich Null ist. Sei etwa $k_j \neq 0$. Dann ist

$$\mathbf{a}_j = \sum_{l \neq j} \left(-\frac{k_l}{k_j} \right) \mathbf{a}_l$$

Man kann sich weiterhin überlegen: Die Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_m$ sind genau dann linear unabhängig, wenn keiner der Vektoren als eine Linearkombination der übrigen Vektoren dargestellt werden kann.

4. Ein Vektor \mathbf{b} wird ein *skalares Vielfaches* eines Vektors \mathbf{a} genannt, wenn es eine Zahl c gibt, so daß $\mathbf{b} = c\mathbf{a}$ ist. Offenbar gilt, daß zwei Vektoren \mathbf{a} und \mathbf{b} genau dann linear unabhängig sind, wenn keiner der beiden ein skalares Vielfaches des jeweils anderen ist.

5. Wenn $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ eine invertierbare (n, n) -Matrix ist, dann sind ihre Spaltenvektoren $\mathbf{a}_1, \dots, \mathbf{a}_n$ linear unabhängig. Denn angenommen, daß $\mathbf{A}\mathbf{k} = \mathbf{0}$ ist, wobei $\mathbf{k} = (k_1, \dots, k_n)'$ irgendein Vektor ist. Dann folgt aus der Invertierbarkeit von \mathbf{A} :

$$\mathbf{k} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{k} = \mathbf{A}^{-1}(\mathbf{A}\mathbf{k}) = \mathbf{0}$$

Es gibt also keine nicht-triviale Linearkombination der Spalten von \mathbf{A} , mit der man einen Nullvektor erzeugen könnte. Es gilt auch umgekehrt: Wenn die $(n, 1)$ -Spaltenvektoren $\mathbf{a}_1, \dots, \mathbf{a}_n$ linear unabhängig sind, dann ist die Matrix

$$\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$$

invertierbar. Ein Beweis ist jedoch etwas komplizierter und soll hier ausgelassen werden.

6. Sei $\mathbf{a}_1, \dots, \mathbf{a}_m$ eine Menge von $(n, 1)$ -Spaltenvektoren. Wenn $m > n$ ist, sind diese Vektoren linear abhängig. Denn angenommen, sie wären linear unabhängig. Dann wären auch die Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_{n+1}$ linear unabhängig, und keiner von ihnen wäre ein Nullvektor. Also bestünde dann auch die Matrix $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ aus linear unabhängigen Spaltenvektoren und wäre invertierbar. Dann könnte man das Gleichungssystem

$$\mathbf{A}\mathbf{k} = \mathbf{a}_1 k_1 + \dots + \mathbf{a}_n k_n = \mathbf{a}_{n+1}$$

betrachten. Da $\mathbf{a}_{n+1} \neq \mathbf{0}$ ist, muß auch $\mathbf{k} \neq \mathbf{0}$ sein. Das aber bedeutet, daß einer der Vektoren, nämlich \mathbf{a}_{n+1} , als eine nicht-triviale Linearkombination von $\mathbf{a}_1, \dots, \mathbf{a}_n$ dargestellt werden kann, was ein Widerspruch zur Annahme der linearen Unabhängigkeit ist.

7. Aus der vorstehenden Überlegung folgt insbesondere:

- Wenn \mathbf{A} eine (n, m) -Matrix und $m > n$ ist, dann sind die Spaltenvektoren von \mathbf{A} linear abhängig.
- Wenn \mathbf{A} eine (n, m) -Matrix ist, dann ist sowohl die Anzahl ihrer linear unabhängigen Spaltenvektoren als auch die Anzahl ihrer linear unabhängigen Zeilenvektoren höchstens gleich $\min\{n, m\}$.

8. Sei \mathbf{A} eine (n, m) -Matrix. Die maximale Anzahl linear unabhängiger Spaltenvektoren von \mathbf{A} wird ihr *Spaltenrang* genannt. Ganz analog wird die maximale Anzahl linear unabhängiger Zeilenvektoren von \mathbf{A} ihr *Zeilenrang* genannt. Oder anders formuliert: Der Zeilenrang von \mathbf{A} ist der Spaltenrang von \mathbf{A}' . Man kann zeigen, daß bei jeder Matrix ihr Zeilenrang und ihr Spaltenrang identisch sind. Man spricht deshalb kurz vom *Rang einer Matrix* und bezeichnet ihn mit $\text{rg}(\mathbf{A})$. Offenbar gilt

$$\text{rg}(\mathbf{A}) \leq \min\{n, m\}$$

wie die zuvor angestellte Überlegung gezeigt hat.

9. Daß bei jeder Matrix Zeilen- und Spaltenrang gleich sind, kann man sich folgendermaßen überlegen (diese Überlegung folgt Searle 1982, S. 169f). Sei \mathbf{A} eine (n, m) -Matrix mit dem Zeilenrang r und dem Spaltenrang s . Zunächst ist klar, daß der Zeilenrang unabhängig von der Anordnung der Spalten und der Spaltenrang unabhängig von der Anordnung der Zeilen in der Matrix \mathbf{A} ist. Man kann auch sicherlich durch Vertauschen von Zeilen und Spalten \mathbf{A} in eine solche Form bringen, daß die ersten r Zeilen und die ersten s Spalten linear unabhängig sind. Diese neue Matrix, die wir \mathbf{A}^* nennen, hat den gleichen Zeilen- und Spaltenrang wie \mathbf{A} und kann also folgendermaßen dargestellt werden:

$$\mathbf{A}^* = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{pmatrix}$$

wobei \mathbf{X} eine (r, s) -Matrix, \mathbf{Y} eine $(r, m-s)$ -Matrix, \mathbf{Z} eine $(n-r, s)$ -Matrix und \mathbf{W} eine $(n-r, m-s)$ -Matrix ist. Die Zeilen von \mathbf{Z} sind nach Voraussetzung linear abhängig von den Zeilen von \mathbf{X} . Also gibt es eine Matrix \mathbf{T} , so daß

$$\mathbf{Z}' = \mathbf{X}'\mathbf{T}' \quad \text{bzw.} \quad \mathbf{Z} = \mathbf{T}\mathbf{X}$$

ist. Jetzt kann man zeigen, daß die Spalten von \mathbf{X} nicht linear abhängig sind. Denn angenommen, sie wären linear abhängig; dann gäbe es einen Vektor $\mathbf{a} \neq \mathbf{0}$, so daß $\mathbf{X}\mathbf{a} = \mathbf{0}$ ist. Dann aber wäre auch $\mathbf{Z}\mathbf{a} = \mathbf{T}\mathbf{X}\mathbf{a} = \mathbf{0}$, also auch:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} \mathbf{a} = \mathbf{0}$$

Das aber wäre ein Widerspruch zur Voraussetzung, daß die ersten s Spalten von \mathbf{A}^* linear unabhängig sind. Also hat \mathbf{X} s linear unabhängige Spalten. Und da, wie bereits gezeigt wurde, die Anzahl der linear unabhängigen Spalten einer Matrix nicht größer als die Anzahl ihrer Zeilen sein kann, folgt $s \leq r$. Mit einer ganz analogen Überlegung kann man zeigen, daß $r \leq s$ gelten muß. Also müssen Zeilen- und Spaltenrang gleich sein.

10. Zwei $(n, 1)$ -Vektoren \mathbf{x} und \mathbf{y} heißen *orthogonal*, wenn $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = 0$ ist. Es gilt: wenn $\mathbf{x} \neq \mathbf{0}$ und $\mathbf{y} \neq \mathbf{0}$ und \mathbf{x} und \mathbf{y} orthogonal sind, dann sind \mathbf{x} und \mathbf{y} auch linear unabhängig. Denn angenommen, sie wären linear abhängig; dann müßte es eine Zahl $a \neq 0$ geben, so daß $\mathbf{x} = a\mathbf{y}$ ist. Da \mathbf{x} und \mathbf{y} orthogonal sind, würde daraus indessen folgen: $\mathbf{x}'\mathbf{x} = a\mathbf{x}'\mathbf{y} = 0$, also auch $\mathbf{x} = \mathbf{0}$, was ein Widerspruch zur Voraussetzung ist.

11. Eine (n, n) -Matrix $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ heißt *orthogonal*, wenn gilt:

$$\mathbf{a}'_j \mathbf{a}_k = \begin{cases} 0 & \text{wenn } j \neq k \\ 1 & \text{wenn } j = k \end{cases} \quad (\text{für alle } j, k = 1, \dots, n)$$

Wenn \mathbf{A} eine orthogonale Matrix ist, sind offenbar alle ihre Spaltenvektoren ungleich Null; und aus der Orthogonalitätsbedingung folgt auch, daß ihre Spaltenvektoren linear unabhängig sind. Eine orthogonale Matrix ist also invertierbar. Weiterhin gilt:

$$\mathbf{A}'\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \mathbf{a}_1 & \cdots & \mathbf{a}'_1 \mathbf{a}_n \\ \vdots & & \vdots \\ \mathbf{a}'_n \mathbf{a}_1 & \cdots & \mathbf{a}'_n \mathbf{a}_n \end{pmatrix} = \mathbf{I}_n$$

Da \mathbf{A} invertierbar ist, folgt:

$$\mathbf{A}' = \mathbf{A}'\mathbf{I}_n = \mathbf{A}'\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n\mathbf{A}^{-1} = \mathbf{A}^{-1}$$

Bei einer orthogonalen Matrix \mathbf{A} ist also ihre Inverse mit \mathbf{A}' identisch.

A.5 Eigenwerte und Eigenvektoren

1. Sei \mathbf{A} eine quadratische (n, n) -Matrix. Ein $(n, 1)$ -Vektor \mathbf{v} heißt ein *Eigenvektor* von \mathbf{A} , wenn $\mathbf{v} \neq \mathbf{0}$ ist und es eine Zahl λ gibt, so daß gilt:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

λ wird dann ein *Eigenwert* von \mathbf{A} genannt. Man sagt auch: λ ist der zum Eigenvektor \mathbf{v} korrespondierende Eigenwert.

2. In diesem Text benötigen wir Eigenwerte und Eigenvektoren nur für symmetrische Matrizen. Dann gibt es zwei grundlegende Feststellungen.² Für jede symmetrische (n, n) -Matrix \mathbf{A} gilt:

- Die Eigenwerte und Eigenvektoren von \mathbf{A} sind stets reell.
- \mathbf{A} hat genau n linear unabhängige Eigenvektoren $\mathbf{v}_1, \dots, \mathbf{v}_n$ mit zugehörigen Eigenwerten $\lambda_1, \dots, \lambda_n$.

²Die Beweise sind ziemlich kompliziert und sollen hier deshalb nicht nachvollzogen werden.

Betrachten wir als Beispiel eine Diagonalmatrix

$$\mathbf{A} = \text{diag}(a_1, \dots, a_n)$$

Sie hat offenbar die Eigenwerte a_1, \dots, a_n und die zugehörigen Eigenvektoren sind gerade die Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$. Dies Beispiel zeigt auch, daß Eigenwerte Null sein können und daß mehrere Eigenwerte einer Matrix gleich sein können.

3. Im weiteren beziehen wir uns auf eine beliebige symmetrische (n, n) -Matrix \mathbf{A} . Es seien $\mathbf{v}_1, \dots, \mathbf{v}_n$ ihre Eigenvektoren und $\lambda_1, \dots, \lambda_n$ die zugehörigen Eigenwerte, so daß man schreiben kann:

$$\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j \quad (\text{für } j = 1, \dots, n) \quad (\text{A.5.1})$$

Man erkennt: Wenn \mathbf{v}_j ein Eigenvektor von \mathbf{A} ist, dann ist auch $k\mathbf{v}_j$ ein Eigenvektor, wobei k eine beliebige Zahl sein kann. Eigenvektoren können also beliebig normiert werden. Insbesondere können sie so normiert werden, daß gilt: $\mathbf{v}'_j \mathbf{v}_j = 1$. Wir werden im folgenden stets diese Normierung voraussetzen.

4. Die Eigenvektoren können zu einer Matrix

$$\mathbf{V} := (\mathbf{v}_1, \dots, \mathbf{v}_n)$$

zusammengefaßt werden. Da die Spalten dieser Matrix linear unabhängig sind, ist es eine invertierbare Matrix. Weiterhin können die Eigenwerte zu einer Diagonalmatrix

$$\mathbf{\Lambda} := \text{diag}(\lambda_1, \dots, \lambda_n)$$

zusammengefaßt werden. Also kann (A.5.1) auch folgendermaßen geschrieben werden (man beachte, daß die Matrizenmultiplikation auf beiden Seiten nicht kommutativ ist):

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda} \quad (\text{A.5.2})$$

Diese Gleichung zeigt auch, daß man die Reihenfolge der Eigenvektoren (Spalten) von \mathbf{V} beliebig verändern kann, wenn man nur die Reihenfolge der Eigenwerte in $\mathbf{\Lambda}$ in entsprechender Weise verändert.

5. Man kann auf einfache Weise zeigen, daß die Eigenvektoren zu zwei verschiedenen Eigenwerten orthogonal sind. Beziehen wir uns auf zwei Eigenwerte λ_j und λ_k . Dann kann man zunächst schreiben:

$$\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{und} \quad \mathbf{A}\mathbf{v}_k = \lambda_k \mathbf{v}_k$$

Aus der ersten Gleichung folgt: $\mathbf{v}'_k \mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}'_k \mathbf{v}_j$. Aus der zweiten Gleichung folgt zunächst, da \mathbf{A} symmetrisch ist: $\mathbf{v}'_k \mathbf{A} = \lambda_k \mathbf{v}'_k$; und daraus

folgt durch Multiplikation mit \mathbf{v}_j : $\mathbf{v}'_k \mathbf{A} \mathbf{v}_j = \lambda_k \mathbf{v}'_k \mathbf{v}_j$. Beide Überlegungen zusammen ergeben also

$$\lambda_j \mathbf{v}'_k \mathbf{v}_j = \lambda_k \mathbf{v}'_k \mathbf{v}_j$$

Wenn $\lambda_j \neq \lambda_k$ ist, folgt daraus $\mathbf{v}'_k \mathbf{v}_j = 0$, d.h., daß die beiden Eigenvektoren orthogonal sind.

6. Aus dieser Überlegung ergibt sich: Wenn alle Eigenwerte einer symmetrischen Matrix \mathbf{A} verschieden sind, besteht die Matrix ihrer Eigenvektoren aus paarweise orthogonalen Spaltenvektoren. Da wir vereinbart hatten, daß die Eigenvektoren normiert sind, d.h. $\mathbf{v}'_j \mathbf{v}_j = 1$, handelt es sich um eine orthogonale Matrix, also

$$\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_n \quad (\text{A.5.3})$$

Die Überlegung, um die Orthogonalität von zwei Eigenvektoren zu zeigen, beruhte allerdings darauf, daß man sich auf zwei verschiedene Eigenwerte beziehen kann. Wenn zwei oder mehr Eigenwerte den gleichen Wert haben, versagt die Überlegung. Man kann aber zeigen, daß auch dann die Eigenvektoren so gewählt werden können, daß sie zueinander orthogonal sind.³ Das heißt, daß die Eigenvektoren einer symmetrischen Matrix immer derart bestimmt werden können, daß die Matrix der Eigenvektoren orthogonal ist, also (A.5.3) gilt.

7. Eine symmetrische (n, n) -Matrix \mathbf{A} hat genau dann den Rang n (und ist also invertierbar), wenn alle ihre Eigenwerte ungleich Null sind. Einen Beweis kann man sich leicht überlegen. (a) Nehmen wir zunächst an, daß $\text{rg}(\mathbf{A}) = n$ ist. Dann müssen alle Eigenwerte ungleich Null sein. Denn wäre z.B. $\lambda_k = 0$, wäre auch $\mathbf{A}\mathbf{v}_k = \mathbf{0}$; und dies würde zeigen, daß die Spalten von \mathbf{A} linear abhängig sind. (b) Nehmen wir umgekehrt an, daß alle Eigenwerte von \mathbf{A} ungleich Null sind. Dann müssen die Spalten von \mathbf{A} linear unabhängig sein. Denn angenommen, es gäbe einen Vektor $\mathbf{k} \neq \mathbf{0}$, so daß $\mathbf{A}\mathbf{k} = \mathbf{0}$ ist. Dann würde aus (A.5.2) folgen:

$$\mathbf{V}'\mathbf{A}\mathbf{k} = \mathbf{\Lambda}\mathbf{V}'\mathbf{k} = \mathbf{0}$$

Dies wäre aber ein Widerspruch, da \mathbf{V}' , wie \mathbf{V} , vollen Rang und damit linear unabhängige Spalten hat.

8. Eine symmetrische Matrix \mathbf{A} heißt *positiv semi-definit*, wenn für alle Vektoren $\mathbf{x} \neq \mathbf{0}$ gilt:⁴

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \quad (\text{A.5.4})$$

³Die Überlegung ist etwas komplizierter und soll hier nicht nachvollzogen werden.

⁴Entsprechend heißt die Matrix *positiv definit*, wenn in der folgenden Bedingung ein strenges $>$ gefordert wird.

Die meisten symmetrischen Matrizen, mit denen wir uns in diesem Text beschäftigen, haben diese Eigenschaft. Insbesondere ist jede Matrix, die als ein *Kreuzprodukt*

$$\mathbf{A} = \mathbf{X}'\mathbf{X}$$

dargestellt werden kann, positiv semi-definit; also insbesondere auch Kovarianz- und Korrelationsmatrizen. Denn gilt diese Darstellung, kann man für jeden beliebigen Vektor \mathbf{x} auch einen Vektor $\mathbf{y} = \mathbf{X}\mathbf{x}$ bilden und findet dann:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{X}'\mathbf{X}\mathbf{x} = \mathbf{y}'\mathbf{y} \geq 0$$

9. Wenn eine symmetrische Matrix \mathbf{A} positiv semi-definit ist, folgt daraus, daß alle ihre Eigenwerte größer oder gleich Null sind. Denn geht man von der Darstellung (A.5.1) aus, folgt

$$\mathbf{v}'_j \mathbf{A} \mathbf{v}_j = \lambda_j \mathbf{v}'_j \mathbf{v}_j \geq 0$$

Also muß gelten, daß $\lambda_j \geq 0$ ist. Entsprechend kann man sich überlegen, daß alle Eigenwerte positiv sind, wenn die Matrix positiv definit ist.

10. Wenn \mathbf{A} positiv semi-definit ist, kann man mit ihren Eigenwerten und Eigenvektoren eine sehr einfache Darstellung finden. Es gibt dann nämlich keine negativen Eigenwerte, und man kann eine Diagonalmatrix

$$\mathbf{\Lambda}^{1/2} := \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$$

definieren. Dann findet man ausgehend von (A.5.2) und (A.5.3):

$$\mathbf{A} = \mathbf{A}\mathbf{V}\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}' = (\mathbf{V}\mathbf{\Lambda}^{1/2})(\mathbf{V}\mathbf{\Lambda}^{1/2})'$$

also $\mathbf{A} = \mathbf{W}\mathbf{W}'$, wobei $\mathbf{W} = \mathbf{V}\mathbf{\Lambda}^{1/2}$ ist, also auch $\mathbf{W}'\mathbf{W} = \mathbf{\Lambda}$.

11. Wenn \mathbf{A} vollen Rang hat, also invertierbar ist, kann mit Hilfe dieser Darstellung die Inverse von \mathbf{A} konstruiert werden. Es kann dann nämlich, wie oben gezeigt worden ist, kein Eigenwert gleich Null, sondern alle Eigenwerte müssen größer als Null sein. Also kann man eine Matrix

$$\mathbf{\Lambda}^{-1} := \begin{pmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_n \end{pmatrix}$$

definieren, und mit ihrer Hilfe auch $\mathbf{A}^{-1} := \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}'$. Einfaches Ausrechnen zeigt, daß es sich um die Inverse von \mathbf{A} handelt.

A.6 Lineare Regression

1. Gelegentlich benötigen wir in diesem Text Verfahren der linearen Regression, wie sie aus Einführungen in die Statistik bekannt sind. Hier soll deshalb gezeigt werden, wie mit Hilfe der Matrixschreibweise eine einfache Darstellung erreicht werden kann. Als Ausgangspunkt nehmen wir an, daß eine abhängige Variable Y und unabhängige Variablen X_2, \dots, X_m gegeben sind. Ein linearer Regressionsansatz sieht dann folgendermaßen aus:

$$Y = \beta_1 + X_2\beta_2 + \dots + X_m\beta_m + U(\beta_1, \dots, \beta_m)$$

wobei die Variable $U(\beta_1, \dots, \beta_m)$ durch den Regressionsansatz definiert wird, so daß ihre Werte auch von den Regressionsparametern β_1, \dots, β_m abhängen. Definiert man eine Hilfsvariable X_1 , die stets den Wert 1 annimmt, kann man auch schreiben:

$$Y = X_1\beta_1 + X_2\beta_2 + \dots + X_m\beta_m + U(\beta_1, \dots, \beta_m) \quad (\text{A.6.1})$$

Die Idee ist, Werte für die Parameter β_1, \dots, β_m so zu bestimmen, daß die Werte der durch den Regressionsansatz definierten Variablen $U(\beta_1, \dots, \beta_m)$ insgesamt „möglichst klein“ werden. Das meistens verwendete Kriterium bezieht sich auf die Summe der quadrierten Werte von $U(\beta_1, \dots, \beta_m)$. Man spricht dann von einer *OLS-Regression* („ordinary least squares“).

2. Um diese Idee praktisch zu verfolgen, benötigt man Daten. Nehmen wir also an, daß uns Werte für die Variablen Y und X_1, \dots, X_m zur Verfügung stehen. Gibt es n Werte, können sie mit Hilfe der Matrixschreibweise so dargestellt werden:

$$\mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{und} \quad \mathbf{X} := \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

Definiert man außerdem einen Vektor $\boldsymbol{\beta} := (\beta_1, \dots, \beta_m)'$ für die Parameter und einen Vektor $\mathbf{u}(\boldsymbol{\beta}) := (u_1(\boldsymbol{\beta}), \dots, u_n(\boldsymbol{\beta}))'$ für die Werte der Variablen $U(\beta_1, \dots, \beta_m)$, kann der Regressionsansatz (A.6.1) in Matrixschreibweise so formuliert werden:

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \dots + \mathbf{x}_m\beta_m + \mathbf{u}(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}(\boldsymbol{\beta}) \quad (\text{A.6.2})$$

wobei $\mathbf{x}_1, \dots, \mathbf{x}_m$ die Spalten der Matrix \mathbf{X} sind. $\mathbf{X}\boldsymbol{\beta}$ ist also eine Linearkombination der Spalten von \mathbf{X} .

3. Um Werte für die Parameter des Regressionsansatzes zu bestimmen, muß der folgende Ausdruck minimiert werden:

$$\sum_{i=1}^n u_i(\boldsymbol{\beta})^2 = \mathbf{u}(\boldsymbol{\beta})'\mathbf{u}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad (\text{A.6.3})$$

Es muß also das Minimum der Funktion

$$f(\boldsymbol{\beta}) := (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

bestimmt werden. Eine notwendige Bedingung ist, daß an der Stelle des Minimums die ersten Ableitungen der Funktion Null werden. In der Matrixschreibweise können diese Ableitungen so ausgedrückt werden:⁵

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = -2\mathbf{X}'\mathbf{u}(\boldsymbol{\beta})$$

Parameterwerte, die ein Minimum der Funktion f liefern, müssen also die Bedingung

$$\mathbf{X}'\mathbf{u}(\boldsymbol{\beta}) = \mathbf{0} \quad \text{bzw.} \quad \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \quad (\text{A.6.4})$$

erfüllen. Offenbar gibt es genau dann eine Lösung, wenn $\mathbf{X}'\mathbf{X}$ eine invertierbare Matrix ist, und die eindeutige Lösung ist dann

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

4. Definiert man $\hat{\mathbf{y}} := \mathbf{X}\hat{\boldsymbol{\beta}}$ und $\mathbf{u}(\hat{\boldsymbol{\beta}}) := \mathbf{y} - \hat{\mathbf{y}}$, erhält man die Darstellung

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{u}(\hat{\boldsymbol{\beta}})$$

Unmittelbar aus (A.6.4) folgt außerdem $\mathbf{X}'\mathbf{u}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, so daß der Vektor $\mathbf{u}(\hat{\boldsymbol{\beta}})$ zu allen Spalten von \mathbf{X} , und somit auch zu jeder Linearkombination dieser Spalten, orthogonal ist.

A.7 Pyrrhons Lemma

1. Um mit der Regressionsrechnung etwas vertrauter zu werden, besprechen wir *Pyrrhons Lemma*, das zeigt, wie man durch Hinzufügen von Variablen zu einem Regressionsansatz die Modellparameter verändern kann. Wir folgen einer Darstellung durch T. K. Dijkstra (1995), der auch die Bezeichnung *Pyrrhons Lemma* vorgeschlagen hat, wohl um zum Ausdruck zu bringen, daß man der Regressionsrechnung mit Skepsis begegnen sollte.

⁵Faßt man $\mathbf{A}\mathbf{x}$ als eine vektorwertige Funktion des Vektors \mathbf{x} auf, gilt folgende Differentiationsregel:

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}'$$

Entsprechend gilt, wenn man $\mathbf{x}'\mathbf{A}\mathbf{x}$ als eine skalarwertige Funktion des Vektors \mathbf{x} auf, die Regel

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

2. Als Ausgangspunkt dient eine lineare Regression der Form

$$\mathbf{y} = \mathbf{x}_1 \hat{\beta}_1 + \cdots + \mathbf{x}_m \hat{\beta}_m + \mathbf{u}(\hat{\beta}) \quad (\text{A.7.1})$$

wobei vorausgesetzt wird, daß die Länge der Vektoren gleich $n > m$ ist und die Regressionsparameter $\hat{\beta}_1, \dots, \hat{\beta}_m$ mit der OLS-Methode berechnet worden sind. Sei jetzt angenommen, daß man sich anstelle der zunächst gefundenen Modellparameter $\hat{\beta}_1, \dots, \hat{\beta}_m$ andere Werte wünscht, etwa $\hat{\beta}_1^*, \dots, \hat{\beta}_m^*$. Pyrrhons Lemma besagt, daß man dies dadurch erreichen kann, daß man eine weitere Variable, wir nennen sie \mathbf{z} , konstruiert und in den Regressionsansatz mit aufnimmt:

$$\mathbf{y} = \mathbf{x}_1 \hat{\beta}_1^* + \cdots + \mathbf{x}_m \hat{\beta}_m^* + \mathbf{z} \hat{\gamma}^* + \mathbf{u}(\hat{\beta}^*, \hat{\gamma}^*)$$

Um das zu erreichen, geht man von folgender Darstellung für die neue Variable \mathbf{z} aus:

$$\mathbf{z} = \mathbf{x}_1 \alpha_1 + \cdots + \mathbf{x}_m \alpha_m + \mathbf{u}(\hat{\beta}) \delta + \mathbf{d} \quad (\text{A.7.2})$$

wobei $\delta \neq 0$ irgendeine Zahl ist und \mathbf{d} irgendein Vektor der Länge n , der sowohl zu $\mathbf{x}_1, \dots, \mathbf{x}_m$ als auch zu $\mathbf{u}(\hat{\beta})$ orthogonal ist, so daß gilt:

$$\mathbf{d}' \mathbf{u}(\hat{\beta}) = 0, \quad \mathbf{d}' \mathbf{x}_1 = 0, \dots, \mathbf{d}' \mathbf{x}_m = 0$$

Wenn $n > m + 1$ ist, kann ein solcher Vektor stets gefunden werden.

3. Verwendet man in dem erweiterten Regressionsansatz diese Darstellung für \mathbf{z} , findet man:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}_1 \hat{\beta}_1^* + \cdots + \mathbf{x}_m \hat{\beta}_m^* + \mathbf{z} \hat{\gamma}^* + \mathbf{u}(\hat{\beta}^*, \hat{\gamma}^*) \\ &= \mathbf{x}_1 (\hat{\beta}_1^* + \alpha_1 \hat{\gamma}^*) + \cdots + \mathbf{x}_m (\hat{\beta}_m^* + \alpha_m \hat{\gamma}^*) + \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*) \end{aligned} \quad (\text{A.7.3})$$

wobei $\mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*) := (\mathbf{u}(\hat{\beta}) \delta + \mathbf{d}) \hat{\gamma}^* + \mathbf{u}(\hat{\beta}^*, \hat{\gamma}^*)$ ist. Faßt man die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$ zu einer Matrix \mathbf{X} zusammen und bildet man den Vektor

$$\mathbf{w} := (\hat{\beta}_1 - (\hat{\beta}_1^* + \alpha_1 \hat{\gamma}^*), \dots, \hat{\beta}_m - (\hat{\beta}_m^* + \alpha_m \hat{\gamma}^*))'$$

kann man die Gleichungen (A.7.1) und (A.7.3) folgendermaßen zusammenfassen:

$$\mathbf{X} \mathbf{w} = \mathbf{u}(\hat{\beta}) - \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*)$$

Also gilt auch

$$\mathbf{X}' \mathbf{X} \mathbf{w} = \mathbf{X}' (\mathbf{u}(\hat{\beta}) - \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*))$$

Nun sind jedoch nicht nur $\mathbf{u}(\hat{\beta})$ und $\mathbf{u}(\hat{\beta}^*, \hat{\gamma}^*)$ orthogonal zu den Regressorvariablen $\mathbf{x}_1, \dots, \mathbf{x}_m$, sondern dies gilt nach Voraussetzung ebenfalls für den Vektor \mathbf{d} . Das gleiche gilt somit auch für $\mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*)$, und es folgt

$$\mathbf{X}' (\mathbf{u}(\hat{\beta}) - \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*)) = \mathbf{0}$$

Also ist auch

$$\mathbf{w} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{u}(\hat{\beta}) - \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*)) = \mathbf{0}$$

und man findet:

$$\mathbf{u}(\hat{\beta}) = \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*) \quad \text{und} \quad \hat{\beta}_j = \hat{\beta}_j^* + \alpha_j \hat{\gamma}^* \quad (\text{A.7.4})$$

für $j = 1, \dots, m$. Hieraus läßt sich zunächst ein beliebiger Wert für $\hat{\gamma}^*$ bestimmen. Denn aus $\mathbf{u}(\hat{\beta}) = \mathbf{u}(\alpha, \hat{\beta}^*, \hat{\gamma}^*)$ folgt zunächst

$$\mathbf{u}(\hat{\beta}) = (\mathbf{u}(\hat{\beta}) \delta + \mathbf{d}) \hat{\gamma}^* + \mathbf{u}(\hat{\beta}^*, \hat{\gamma}^*)$$

Multipliziert man diese Gleichung von links mit $(\mathbf{u}(\hat{\beta}) \delta + \mathbf{d})'$, und berücksichtigt man, daß aus den Orthogonalitätsbeziehungen insbesondere auch $(\mathbf{u}(\hat{\beta}) \delta + \mathbf{d})' \mathbf{u}(\hat{\beta}^*, \hat{\gamma}^*) = \mathbf{0}$ folgt, findet man:

$$\mathbf{u}(\hat{\beta})' \mathbf{u}(\hat{\beta}) \delta = (\mathbf{u}(\hat{\beta}) \delta + \mathbf{d})' (\mathbf{u}(\hat{\beta}) \delta + \mathbf{d}) \hat{\gamma}^* = (\mathbf{u}(\hat{\beta})' \mathbf{u}(\hat{\beta}) \delta^2 + \mathbf{d}' \mathbf{d}) \hat{\gamma}^*$$

Und daraus gewinnt man schließlich die Darstellung

$$\hat{\gamma}^* = \frac{\mathbf{u}(\hat{\beta})' \mathbf{u}(\hat{\beta}) \delta}{\mathbf{u}(\hat{\beta})' \mathbf{u}(\hat{\beta}) \delta^2 + \mathbf{d}' \mathbf{d}} \quad (\text{A.7.5})$$

Beginnt man also mit den Residuen $\mathbf{u}(\hat{\beta})$ des ursprünglichen Regressionsansatzes und einem in weiten Grenzen frei wählbaren Vektor \mathbf{d} , kann man sich durch die Wahl der Zahl δ zunächst einen beliebigen Koeffizienten $\hat{\gamma}^*$ verschaffen. Dann aber kann man ausgehend von beliebig vorgegebenen neuen Modellparametern $\hat{\beta}_1^*, \dots, \hat{\beta}_m^*$ aufgrund von (A.7.4) die erforderlichen Koeffizienten $\alpha_1, \dots, \alpha_m$ bestimmen, um schließlich mit Hilfe von (A.7.2) die Werte der zusätzlichen Regressorvariablen \mathbf{z} zu berechnen.

4. Um den Rechengang zu illustrieren, verwenden wir folgende (willkürlich ausgedachten) Daten:

\mathbf{y}	\mathbf{x}_1	\mathbf{x}_2	\mathbf{y}	\mathbf{x}_1	\mathbf{x}_2
1600	1	1	2900	1	0
2000	1	1	3500	1	0
1800	1	0	2600	1	1
2500	1	1	4000	1	0

Man kann sich vorstellen, daß es sich um Daten für 8 Personen handelt, wobei \mathbf{y} das monatliche Erwerbseinkommen und \mathbf{x}_2 das Geschlecht (0 männlich, 1 weiblich) erfaßt. Der Regressionsansatz ist dann:

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{u}(\beta) \quad (\text{A.7.6})$$

und man erhält die Modellparameter $\hat{\beta}_1 = 3050$ und $\hat{\beta}_2 = -875$, woraus erkenntlich wird, daß Frauen im Durchschnitt weniger verdienen als Männer. Aber angenommen, jemand hätte lieber andere Modellparameter, etwa

$$\hat{\beta}_1^* = 3050 \quad \text{und} \quad \hat{\beta}_2^* = 0$$

so daß die Geschlechtsvariable keinen Beitrag zum bedingten Mittelwert liefert. Pyrrhons Lemma zeigt dann, wie man eine Variable \mathbf{z} konstruieren kann, deren Aufnahme in den Regressionsansatz zu diesem Ergebnis führt. Zunächst muß man irgendeinen Vektor \mathbf{d} finden, der sowohl zu \mathbf{x}_1 und \mathbf{x}_2 als auch zu $\mathbf{u}(\hat{\beta})$ orthogonal ist. Dafür gibt es viele verschiedene Möglichkeiten. Die folgende Tabelle zeigt zunächst in den ersten beiden Spalten noch einmal \mathbf{x}_1 und \mathbf{x}_2 und in der dritten Spalte den Residualvektor $\mathbf{u}(\hat{\beta})$ aus der Regression (A.7.6). Dann folgen in der vierten Spalte Werte eines Vektors \mathbf{d} , der zu \mathbf{x}_1 , \mathbf{x}_2 und $\mathbf{u}(\hat{\beta})$ orthogonal ist.

\mathbf{x}_1	\mathbf{x}_2	$\mathbf{u}(\hat{\beta})$	\mathbf{d}	\mathbf{z}
1	1	-575	-15.25	-161.677
1	1	-175	17.25	-89.177
1	0	-1250	3.00	-122.000
1	1	325	-1.00	-57.427
1	0	-150	-1.00	-16.000
1	0	450	-1.00	44.000
1	1	425	-1.00	-47.426
1	0	950	-1.00	94.000

Somit findet man $\mathbf{u}(\hat{\beta})'\mathbf{u}(\hat{\beta}) = 3337500$ und $\mathbf{d}'\mathbf{d} = 544.125$. Wählt man z.B. $\delta = 0.1$, erhält man aus (A.7.5) den Wert $\hat{\gamma}^* \approx 9.8396$. Also gewinnt man aus den gewünschten neuen Modellparametern und aus (A.7.4) die Werte $\alpha_1 = 0$ und $\alpha_2 = -88.9265$. Und so kann man schließlich die Werte der neuen Variablen \mathbf{z} berechnen, die in der letzten Spalte der obigen Tabelle angegeben sind. Der Regressionsansatz

$$\mathbf{y} = \mathbf{x}_1\beta_1^* + \mathbf{x}_2\beta_2^* + \mathbf{z}\gamma^* + \mathbf{u}(\hat{\beta}^*, \gamma^*)$$

liefert dann die gewünschten neuen Modellparameter, nämlich $\hat{\beta}_1^* \approx 3050$, $\hat{\beta}_2^* \approx 0$ und $\hat{\gamma}^* \approx 9.84$.

Literatur

- Adams, E. W. 1966. On the Nature and Purpose of Measurement. *Synthese* 16, 125–169.
- Adorno, T. W. 1957. Soziologie und empirische Forschung. In: E. Topitsch (Hg.), *Logik der Sozialwissenschaften*, 511–525. Köln: Kiepenheuer & Witsch 1972.
- Ajzen, I. 1988. *Attitudes, Personality, and Behavior*. Stony Stratford: Open University Press.
- Alber, J. 1988. Die Gesundheitssysteme der OECD-Länder im Vergleich. In: M. G. Schmidt (Hg.), *Staatstätigkeit. Internationale und historisch vergleichende Analysen*, 116–150. Opladen: Westdeutscher Verlag.
- Albrecht, G. 1975. Nicht-reaktive Messung und Anwendung historischer Methoden. In: *Techniken der empirischen Sozialforschung*, Band 2, 9–81. München: Oldenbourg.
- Allmendinger, J., Schmidt, P., Wegener, B. 1983. *ZUMA-Handbuch sozialwissenschaftlicher Skalen. Dokumentarische Bearbeitung*: H. P. Ohly, T. Eikelmann. Bonn: Informationszentrum Sozialwissenschaften.
- Allport, G. W. 1935. *Attitudes*. In: C. Murchison (ed.), *A Handbook of Social Psychology*. Worcester: Clark University Press.
- Anastasi, A. 1982. *Psychological Testing*. 5th ed. New York: Macmillan.
- Andersen, E. B. 1973. *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygiejnisk Forlag.
- Anderson, N. H. 1982. *Cognitive Algebra and Social Psychophysics*. In: B. Wegener (ed.), *Social Attitudes and Psychophysical Measurement*, 123–148. Hillsdale: Lawrence Earlbaum.
- Andrich, D. 1978. A Rating Formulation for Ordered Response Categories. *Psychometrika* 43, 561–573.
- Arminger, G. 1979. *Faktorenanalyse*. Stuttgart: Teubner.
- Atteslander, P. 1975. *Methoden der empirischen Sozialforschung*. Berlin: de Gruyter.
- Atteslander, P., Kopp, M. 1984. Befragung. In: E. Roth, K. Heidenreich (Hg.), *Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis*, 144–172. München: Oldenbourg.
- Backhaus, K., Erichson, B., Plinke, W., Schuchard-Fischer, C., Weiber, R. 1987. *Multivariate Analysemethoden*. Berlin: Springer-Verlag.
- Bailey, K. D. 1982. *Methods of Social Research*. 2nd ed. New York: Free Press.
- Bateson, N. 1984. *Data Construction in Social Surveys*. London: Allen & Unwin.
- Bender, S., Hilzendegen, J., Rohwer, G., Rudolph, H. 1996. Die IAB-Beschäftigtenstichprobe 1975–1990. *BeitrAB 197*. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Besozzi, C., Zehnpfennig, H. 1976. Methodologische Probleme der Index-Bildung. In: *Techniken der empirischen Sozialforschung*, Band 5, 9–55. München: Oldenbourg.
- Birchfield, V., Crepez, M. M. L. 1998. The Impact of Constitutional Structures and Collective and Competitive Veto Points on Income Inequality in Industrialized Democracies. *European Journal of Political Research* 34, 175–200.