

# Kapitel 1

## Multivariate Regression

Gerhard Tutz, Skript LMU, 7. Mai 2014

1.1	Grundkonzept . . . . .	3
1.2	Das klassische multivariate lineare Modell . . . . .	6
1.2.1	Modelle . . . . .	6
1.2.2	Kleinste-Quadrate-Schätzung der Parameter . . . . .	7
1.2.3	Schätzung der Varianz . . . . .	8
1.2.4	Motivation des linearen Regressionsmodells . . . . .	8
1.3	Klassisches lineares Modell in Matrixdarstellung . . . . .	9
1.3.1	Parameterschätzung in Matrixdarstellung . . . . .	10
1.3.2	Eigenschaften der Schätzung und Konfidenzintervalle . . . . .	11
1.3.3	Tests . . . . .	12
1.4	Vorteile des multivariaten linearen Modells . . . . .	15
1.5	Generellere Modelle . . . . .	16
1.5.1	Generelles lineares Modell . . . . .	16
1.5.2	Robuste Schätzung durch generalisierte Schätzgleichungen . . . . .	17

### 1.1 Grundkonzept

In der multivariaten Regression wird wie in jeder Regressionsmodellierung versucht, Variabilität einer abhängigen Größe durch Einflußgrößen zu erklären. Problemstellungen sind insbesondere

- Formulierung eines plausiblen Modells für die Wirkung der Einflußgrößen auf die abhängige Größe,

- Quantifizierung der Wirkung von Einflußgrößen,
- Bestimmung der statistischen Signifikanz von Effekten,
- Prädiktion der abhängigen Größe bei neuen Beobachtungen.

In der Regression haben sich viele Synonyme herausgebildet, die abhängige Größe wird auch als *Regressand*, *zu erklärende Variable* oder *endogene Variable* bezeichnet, während die Einflußgrößen als *Regressoren*, *erklärende Variablen*, *exogene Variablen* oder *Prädiktoren* bezeichnet werden. Üblicherweise bezeichnet für  $i = 1, \dots, n$

$\mathbf{x}_i$  den Vektor der Einflußgrößen und  
 $\mathbf{y}_i$  den Vektor der abhängigen Größen.

Man spricht von *multipler Regression*, wenn zwar  $\mathbf{x}_i$  vektoriell, aber  $y_i$  eindimensional ist. Echt *multivariate Regression* liegt vor, wenn auch  $\mathbf{y}_i$  vektoriell ist.

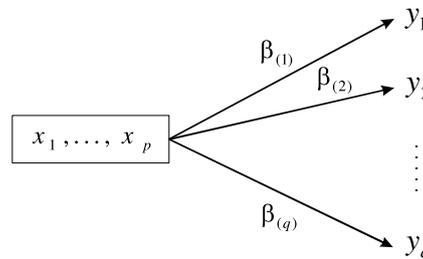


ABBILDUNG 1.1: Multivariate Regression

Multivariate Regressionsmodellierung ist immer dann angebracht, wenn zu Einflußgrößen  $\mathbf{x}_i$  eine ganze Reihe von Messungen vorliegt, die als abhängig zu betrachten sind und von denen zu erwarten ist, daß sie korreliert sind. Ein wichtiger Anwendungsbereich sind *Meßwiederholungen* als spezieller Fall von *Longitudinaldaten*. Wird ein Merkmal wiederholt gemessen, erhält man unmittelbar eine Meßreihe. Das kann im medizinischen Bereich der Cholesterinspiegel über eine Woche sein.

Cholesterin (Tag 1), ... Cholesterin (Tag q),

in Abhängigkeit von Medikamentendosierung, Alter, Geschlecht und weiteren Kovariablen, oder im ökonomischen Bereich der Umsatz über einen gewissen Zeitraum

Umsatz (Woche 1), ... Umsatz (Woche q),

in Abhängigkeit von Wertemaßnahmen. Multivariate Regression ist aber auch dann angebracht, wenn die abhängigen Größen unterschiedliche Merkmale repräsentieren.

In einer Untersuchung zur Schadstoffbelastung können unterschiedliche Schadstoffkonzentrationen

$$\text{Schadstoff}_1, \dots \quad \text{Schadstoff}_q,$$

in Abhängigkeit von Windgeschwindigkeit, Temperatur, Standort von Interesse sein. Im biometrischen Bereich können die abhängigen Merkmale durch unterschiedliche Körpermaße

$$\text{Gewicht, Größe, } \dots \quad \text{Brustumfang,}$$

gemessen an einer Person in Abhängigkeit von Körpermaßen gegeben sein.

Allen diesen Messungen ist gemeinsam, daß von einer Korrelation zwischen den abhängigen Größen auszugehen ist. Die Grundstruktur der Abhängigkeit ist in Abbildung 1.1 dargestellt. Ein Vektor von unabhängigen Größen  $\mathbf{x}^T = (x_1, \dots, x_p)$  wirkt auf die Messungen  $y_1, \dots, y_q$ .

Bei Regressionsmodellen lassen sich zwei Komponenten unterscheiden, die strukturelle oder systematische Komponente und die zufällige Komponente. Die *strukturelle Komponente* spezifiziert, wie der Erwartungswert der abhängigen Variable, die eigentliche Regressionsfunktion, von den Einflußgrößen abhängt. Die *zufällige Komponente* spezifiziert die Fehlerstruktur, d.h. die Verteilung der abhängigen Größe um ihren Erwartungswert. Als abhängige Größe ist hier immer der gesamte Vektor abhängiger Variablen  $\mathbf{y}_i^T = (y_{i1}, \dots, y_{iq})$  zu verstehen.

Die im folgenden betrachteten linearen multivariaten Regressionsmodelle haben die folgende Grundstruktur.

Lineares multivariates Modell
<i>Strukturkomponente</i>
$E(\mathbf{y}_i   \mathbf{x}_i) = \mathbf{X}_i \boldsymbol{\beta}$
für geeignete Designmatrix $\mathbf{X}_i$ und Parametervektor $\boldsymbol{\beta}$
<i>Zufällige Komponente/Fehlerstruktur</i>
$\mathbf{y}_i   \mathbf{x}_i$ folgt einer vorgegebenen Verteilung,
z.B. $\mathbf{y}_i   \mathbf{x}_i$ ist normalverteilt mit Kovarianzmatrix $\boldsymbol{\Sigma}$

Für die Beobachtungen  $(\mathbf{y}_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$ , wird darüberhinaus vorausgesetzt, daß  $\mathbf{y}_1, \dots, \mathbf{y}_n$  bei gegebenen Einflußgrößen  $\mathbf{x}_1, \dots, \mathbf{x}_n$  unabhängig sind.

Stochastische und deterministische Einflußgrößen

## 1.2 Das klassische multivariate lineare Modell

### 1.2.1 Modelle

Zum  $(p + 1)$ -dimensionalen Einflußgrößenvektor  $\mathbf{x}_i^T = (x_{i0}, x_{i1}, \dots, x_{ip})$  betrachtet man den  $q$ -dimensionalen abhängigen Zufallsvektor  $\mathbf{y}_i^T = (y_{i1}, \dots, y_{iq})$  (bei gegebenem  $\mathbf{x}_i$ ),  $i = 1, \dots, n$ . Die erste Komponente des Einflußgrößenvektors ist meist als Konstante festgelegt,  $x_{i0} = 1, i = 1, \dots, n$ , um einen Achsenabschnitt zu modellieren.

Im linearen Modell wird für die *strukturelle Komponente* angenommen, daß der Einflußgrößenvektor  $\mathbf{x}_i$  die  $j$ te Komponente der abhängigen Variable linear bestimmt

$$E(y_{ij}|\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_{(j)} = \beta_{0j} + x_{i1}\beta_{1j} + \dots + x_{ip}\beta_{pj}. \quad (1.1)$$

Die Wirkung auf die  $j$ te Komponente ist somit durch den für die Komponente spezifischen Parametervektor

$$\boldsymbol{\beta}_{(j)}^T = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})$$

bestimmt. Man nimmt damit für jede Komponente der abhängigen Variablen eine lineare Wirkungsstruktur an, wobei der Gewichtsvektor spezifisch für diese Komponente ist. Für die  $j$ te Komponente der abhängigen Variablen ist der Gewichtsvektor  $\boldsymbol{\beta}_{(j)}$  wirksam. Da (1.1) ein lineares Regressionsmodell ist, bleibt damit die Interpretation der Regressionskoeffizienten des multiplen Modells erhalten. Man muß nur berücksichtigen, daß die Wirkung spezifisch für eine Komponente ist. Die Komponenten des Modells (1.1) lassen sich auf unterschiedliche Art und Weise zusammenfassen. Eine Darstellung beruht darauf, daß man alle Messungen an einer statistischen Einheit  $y_{i1}, \dots, y_{iq}$  zu einem Vektor zusammenfaßt. In dieser "Beobachtungsdarstellung" erhält man für die  $i$ te Messung das Modell für den Erwartungswert  $\boldsymbol{\mu}_i = E(\mathbf{y}_i|\mathbf{x}_i)$  in Matrixschreibweise durch

$$\boldsymbol{\mu}_i = \begin{pmatrix} E(y_{i1}|\mathbf{x}_i) \\ \vdots \\ E(y_{iq}|\mathbf{x}_i) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^T & & \\ & \ddots & \\ & & \mathbf{x}_i^T \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{(1)} \\ \vdots \\ \boldsymbol{\beta}_{(q)} \end{pmatrix}$$

bzw. kurz

$$\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta},$$

wobei  $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_{(1)}^T, \dots, \boldsymbol{\beta}_{(q)}^T)$  alle Parametervektoren für die einzelnen Komponenten von  $\mathbf{y}$  zusammenfaßt und  $\boldsymbol{\mu}_i$  den Erwartungswertvektor bezeichnet.

Für die *zufällige Komponente* spezifiziert man

(V)  $\mathbf{y}_i$  besitzt die Kovarianz  $\boldsymbol{\Sigma}$ .

Damit wird eine Form der *Varianzhomogenität* postuliert, d.h. die Kovarianz von  $\mathbf{y}_i|\mathbf{x}_i$  hängt nicht von  $i$  und damit nicht von den Einflußgrößen ab. Jede Meßreihe  $\mathbf{y}_i$  besitzt dieselbe Kovarianzstruktur. Häufig nimmt man zusätzlich Normalverteilung an und postuliert spezifischer

(N)  $\mathbf{y}_i$  ist normalverteilt mit Kovarianzmatrix  $\Sigma$ .

Zusätzlich postuliert man die Unkorreliertheit der Beobachtungsvektoren

$$\text{cov}(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{0} \text{ für } i \neq j$$

Das Modell ist damit gegeben in der folgenden Form

Lineares Normalverteilungsmodell	
<i>Strukturkomponente</i>	$\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$
<i>Verteilungskomponente</i>	$\text{cov}(\mathbf{y}_i) = \Sigma$ (Varianzhomogenität),
	$\mathbf{y}_i$ ist normalverteilt, $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \Sigma_{\mathbf{y}})$

Faßt man alle Beobachtungsvektoren zusammen zu  $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)$  erhält man

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \Sigma_{\mathbf{y}}),$$

wobei

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \quad \text{cov}(\mathbf{y}) = \Sigma_{\mathbf{y}} = \begin{pmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{pmatrix}.$$

### 1.2.2 Kleinste-Quadrate-Schätzung der Parameter

Bei der Kleinsten-Quadrate Methode werden die quadrierten Abstände zwischen den Beobachtungen  $\mathbf{y}_i$  und den dafür durch das Modell angenommenen Erwartungswerten  $\mathbf{X}_i \boldsymbol{\beta}$  in Abhängigkeit von  $\boldsymbol{\beta}$  minimiert. Man betrachtet das Kleinste-Quadrate (least-Squares-) Kriterium

$$LS(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^q (y_{ij} - \mathbf{x}_i^T \boldsymbol{\beta}_{(j)})^2 = \sum_{i=1}^n \|\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}\|^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

wobei  $\|\cdot\|$  die Norm oder Länge eines Vektors bezeichnet ( $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$ ). Einfache Matrixalgebra liefert (die Invertierbarkeit von  $\mathbf{X}^T \mathbf{X}$  vorausgesetzt) die aus der üblichen multiplen Regression vertraute Form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (1.2)$$

Die Kleinste-Quadrate-Schätzung ignoriert die Korreliertheit innerhalb der Beobachtungsvektoren. Anstatt dem einfachen KQ-Kriterium

$$LS(\boldsymbol{\beta}) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$$

scheint das gewichtete Kleinste-Quadrat-Kriterium

$$WLS(\boldsymbol{\beta}) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$$

adäquater. Es läßt sich jedoch zeigen, daß für die zugrundeliegenden speziellen Designmatrizen  $\mathbf{X}_i$  die Maximierung von  $WLS(\boldsymbol{\beta})$  dasselbe Ergebnis liefert, wie die Maximierung von  $LS(\boldsymbol{\beta})$ . Der Kleinste-Quadrate-Schätzer ist daher äquivalent zum gewichteten Kleinste-Quadrate-Schätzer (siehe auch Mardia et al., 1979, Section 6.6.3). Darüberhinaus ist der Kleinste-Quadrate-Schätzer äquivalent zum Maximum-Likelihood-Schätzer unter der Annahme einer multivariaten Normalverteilung.

### 1.2.3 Schätzung der Varianz

Die Schätzung für die Kovarianzmatrix geht von den Residuen aus, die bestimmt sind durch

$$\hat{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}.$$

Für die Differenz zwischen  $\mathbf{y}_i$  und dem wahren  $\mathbf{X}_i \boldsymbol{\beta}$  gilt offensichtlich  $\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ . Wie bei der Schätzung der empirischen Kovarianzmatrix benutzt man daher den Schätzer

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-p-1} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T.$$

Die Normierung ist so gewählt, daß  $\hat{\boldsymbol{\Sigma}}$  ein erwartungstreuer Schätzer ist.

### 1.2.4 Motivation des linearen Regressionsmodells

Das lineare Modell läßt sich ableiten als die beste Prognose für *gemeinsam normalverteilte* Merkmale. Dazu betrachtet man die beiden Zufallsvektoren  $\mathbf{y}^T = (y_1, \dots, y_q)$ ,  $\mathbf{x}^T = (x_1, \dots, x_p)$ .

Unter der *besten Prognose von y* versteht man diejenige von  $\mathbf{x}$  abhängende Prognose  $\hat{\mathbf{y}}(\mathbf{x})$ , für die

$$E(\|\mathbf{y} - \hat{\mathbf{y}}(\mathbf{x})\|^2) = E((\mathbf{y} - \hat{\mathbf{y}}(\mathbf{x}))^T (\mathbf{y} - \hat{\mathbf{y}}(\mathbf{x})))$$

minimal wird. Die Maximierung erfolgt dabei unter allen q-dimensionalen Transformationen von  $\mathbf{x}$ , deren erste und zweite Momente existieren. Als Lösung ergibt sich die bedingte Erwartung  $E(\mathbf{y}|\mathbf{x})$  unter der Bedingung  $\mathbf{x}$ .  $E(\mathbf{y}|\mathbf{x})$  heißt auch Regressionsfunktion erster Art.

Unter der Annahme, daß  $(\mathbf{y}, \mathbf{x})$  gemeinsam normalverteilt sind mit

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \boldsymbol{\Sigma}_{(y,x)}\right), \quad \boldsymbol{\Sigma}_{(y,x)} = \begin{pmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x \end{pmatrix}.$$

ergibt sich, daß die bedingte Erwartung normalverteilt ist mit Erwartungswert  $\mu_{y|x}$  und Kovarianzmatrix  $\boldsymbol{\Sigma}_{y|x}$  in der Form

$$\begin{aligned} \mu_{y|x} &= \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}_x), \\ \boldsymbol{\Sigma}_{y|x} &= \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{xy}. \end{aligned}$$

Daraus folgt, daß die bedingte Erwartung von  $y|x$  linear in  $\mathbf{x}$  ist,

$$\mu_{y|x} = \mathbf{b}_0 + \mathbf{B}_x \mathbf{x},$$

wobei  $\mathbf{b}_0 = \boldsymbol{\mu}_y - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\mu}_x$ ,  $\mathbf{B}_x = \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1}$ . Nimmt man mit  $\mathbf{x}_1^T = (1, \mathbf{x}^T)$  den konstanten Term hinzu, erhält man

$$\mu_{y|x} = \mathbf{B} \mathbf{x}_1, \quad \mathbf{B} = (\mathbf{1}, \mathbf{B}_x).$$

Das entspricht der im linearen Regressionsmodell zugrundegelegten Einflußgrößenstruktur für die die  $j$ te Komponente der abhängigen Größe bestimmt ist durch

$$\mu_{y_j|x} = \mathbf{x}_1^T \boldsymbol{\beta}_{(j)},$$

wobei  $\boldsymbol{\beta}_{(j)}$  die  $j$ te Zeile von  $\mathbf{B}$  bezeichnet. Darüberhinaus gilt die Varianzhomogenität, da  $\boldsymbol{\Sigma}_{y|x}$  nicht von der Größe  $\mathbf{x}$  abhängt, d.h. keine Funktion von  $\mathbf{x}$  ist. Man beachte, daß  $\mathbf{x}$  auf der rechten Seite der Gleichung  $\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{xy}$  nur als Index, nicht als Argument vorkommt.

### 1.3 Klassisches lineares Modell in Matrixdarstellung

Oft wird das klassische multivariate Modell in anderer Struktur dargestellt. Der erste Unterschied zur Darstellung in Abschnitt 1.2 besteht darin, daß man statt der Trennung in Strukturkomponente und zufällige Komponente formuliert

$$y_{ij} = \mathbf{x}_i^T \boldsymbol{\beta}_{(j)} + \epsilon_{ij}, \quad (1.3)$$

wobei  $\epsilon_{ij}$  die Störgröße darstellt, für die  $E(\epsilon_{ij}) = 0$  gilt. Damit ergibt sich wiederum  $\mu_{ij} = E(y_{ij} | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_{(j)}$ .

Der zweite Unterschied besteht in der Anordnung der Beobachtungen. Anstatt den  $i$ ten Beobachtungsvektor als Spaltenvektor zu betrachten wird er als Zeilenvektor dargestellt. Aus (1.3) erhält man unmittelbar die *Matrixdarstellung des multivariaten linearen Modells*

$$\begin{pmatrix} y_{11} & \cdots & y_{1q} \\ \vdots & & \\ y_{n1} & \cdots & y_{nq} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \vdots & & & \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pq} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1q} \\ \vdots & & \\ \epsilon_{n1} & \cdots & \epsilon_{nq} \end{pmatrix}$$

kurz

$$\mathbf{Y} = \mathbf{X}_0 \mathbf{B} + \mathbf{E}.$$

Diese Matrixdarstellung läßt sich unterschiedlich strukturiert betrachten. Für die Beobachtungsvektoren  $\mathbf{y}_i$ ,  $\mathbf{x}_i$  ergibt sich

$$\begin{pmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} (\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(q)}) + \begin{pmatrix} \boldsymbol{\epsilon}_1^T \\ \vdots \\ \boldsymbol{\epsilon}_n^T \end{pmatrix},$$

wobei  $\boldsymbol{\epsilon}_i^T = (\epsilon_{i1}, \dots, \epsilon_{iq})$  die Störgrößen zur  $i$ ten Beobachtung darstellen. Strukturiert man nach den Spalten der Matrix  $\mathbf{Y}$  und damit nach den Komponenten von  $\mathbf{y}$ , ergibt sich

$$(\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(q)}) = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} (\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(q)}) + (\boldsymbol{\epsilon}_{(1)}, \dots, \boldsymbol{\epsilon}_{(q)}),$$

wobei  $\mathbf{y}_{(j)}^T = (y_{1j}, \dots, y_{nj})$  und  $\boldsymbol{\epsilon}_{(j)}^T = (\epsilon_{1j}, \dots, \epsilon_{nj})$  die Beobachtungen und Störgrößen der  $j$ ten Komponente zusammenfaßt.

Die Annahmen für die Matrixdarstellung sind  $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ ,  $\text{cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$ ,  $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \mathbf{0}$  (Unkorreliertheit der Beobachtungen) und meist Normalverteilung  $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ .

### 1.3.1 Parameterschätzung in Matrixdarstellung

Bei der Kleinste-Quadrate-Methode minimiert man

$$LS(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^q (y_{ij} - \mathbf{x}_i^T \boldsymbol{\beta}_{(j)})^2 = \sum_{j=1}^q \sum_{i=1}^n (y_{ij} - \mathbf{x}_i^T \boldsymbol{\beta}_{(j)})^2 = \sum_{j=1}^q LS(\boldsymbol{\beta}_{(j)}), \quad (1.4)$$

wobei  $LS(\boldsymbol{\beta}_{(j)}) = \sum_{i=1}^n (y_{ij} - \mathbf{x}_i^T \boldsymbol{\beta}_{(j)})^2$  das Kleinste-Quadrate Problem für die Beobachtungen der  $j$ ten Komponente darstellen. Dies entspricht der separaten Minimierung der *multiplen Modelle*

$$\mathbf{y}_{(j)} = \mathbf{X}_0 \boldsymbol{\beta}_{(j)} + \boldsymbol{\epsilon}_{(j)},$$

mit der bekannten Lösung

$$\hat{\boldsymbol{\beta}}_{(j)} = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{y}_{(j)},$$

wobei  $\mathbf{y}_{(j)}^T = (y_{1j}, \dots, y_{nj})$  alle Beobachtungen der  $j$ ten Komponente enthält und für  $\mathbf{X}_0$  voller Rang vorausgesetzt wird.

Da  $\mathbf{B} = (\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(q)})$  erhält man die Schätzmatrix

$$\hat{\mathbf{B}} = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{Y}.$$

$\hat{\mathbf{B}}$  ist auch Maximum-Likelihood-Schätzer, wenn  $\mathbf{Y}$  als normalverteilt angenommen wird. Weiter läßt sich zeigen, daß  $\hat{\mathbf{B}}$  die Spur  $\text{tr}((\mathbf{Y} - \mathbf{X}_0\mathbf{B})^T(\mathbf{Y} - \mathbf{X}_0\mathbf{B}))$  und die Determinante  $\det((\mathbf{Y} - \mathbf{X}_0\mathbf{B})^T(\mathbf{Y} - \mathbf{X}_0\mathbf{B}))$  minimiert.

Der gesamte Vektor hat damit die Form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{y}_0,$$

wobei  $\mathbf{y}_0$  strukturiert ist durch  $\mathbf{y}_0^T = (\mathbf{y}_{(1)}^T, \dots, \mathbf{y}_{(q)}^T)$  und  $\text{BlockD}\{(\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T\}$  eine Blockdiagonalmatrix ist mit den Blöcken  $(\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T$ .

Die zugehörige Residuenmatrix hat die Form

$$\hat{\mathbf{E}} = \mathbf{Y} - \mathbf{X}_0\mathbf{B} = \mathbf{Y} - \hat{\mathbf{Y}}_0$$

und der Varianzschätzer läßt sich darstellen durch

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \frac{1}{n-p-1} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T \\ &= \frac{1}{n-p-1} \hat{\mathbf{E}}^T \hat{\mathbf{E}}. \end{aligned}$$

Man beachte, daß diese Darstellung des Kleinste-Quadrate-Schätzers sich von der Darstellung in (1.2) durch  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$  unterscheidet. In Darstellung (1.2) sind die Daten in der Form  $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)$  angeordnet, während die Anordnung in  $\mathbf{y}_0$  nach den abhängigen Komponenten strukturiert ist.

### 1.3.2 Eigenschaften der Schätzung und Konfidenzintervalle

Wenn  $\mathbf{X}_0$  vollen Rang besitzt ( $\text{rg} \mathbf{X}_0 = p+1$ ) und  $p+q+1 \leq n$  ergibt sich

(1)  $\hat{\boldsymbol{\beta}}$  und  $\hat{\boldsymbol{\Sigma}}$  sind erwartungstreu,  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ ,  $E(\hat{\boldsymbol{\Sigma}}) = \boldsymbol{\Sigma}$ .

(2) Die Kovarianz und Varianz der Schätzung sind bestimmt durch

$$\text{cov}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(j)}) = \sigma_{ij} (\mathbf{X}_0^T \mathbf{X}_0)^{-1},$$

wobei  $\boldsymbol{\Sigma} = (\sigma_{ij})$ . Insbesondere gilt damit

$$\text{cov}(\hat{\boldsymbol{\beta}}_{(i)}) = \sigma_{ii} (\mathbf{X}_0^T \mathbf{X}_0)^{-1}.$$

(3)  $\hat{\boldsymbol{\beta}}$  ist bester unverzerrter linearer Schätzer (BLUE, best linear unbiased estimator).

Die Aussagen werden auch als Gauß-Markov-Theorem bezeichnet. Sie gelten auch ohne Normalverteilungsvoraussetzung.

Wenn  $\mathbf{y}_i$  normalverteilt ist, genauer  $\mathbf{y}_i | \mathbf{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$  und das multivariate Modell zugrundeliegt, gelten die weiteren Eigenschaften

$$(4) (n - p - 1) \hat{\Sigma} \sim W_q(\Sigma, n - p - 1),$$

(5)  $\hat{\beta}$  und  $\hat{\Sigma}$  sind unabhängig.

$(1 - \alpha)$ -Konfidenzintervalle für individuelle Parameter  $\beta_{ij}$  haben die Form

$$\hat{\beta}_{ij} - t_{1-\alpha/2}(n - p - 1) \sqrt{\hat{\sigma}_{jj} a_{ii}} \leq \beta_{ij} \leq \hat{\beta}_{ij} + t_{1-\alpha/2}(n - p - 1) \sqrt{\hat{\sigma}_{jj} a_{ii}},$$

wobei  $a_{ii}$  das  $i$ te Diagonalelement von  $(\mathbf{X}_0^T \mathbf{X}_0)^{-1}$  bezeichnet. Sie leiten sich ab aus den Eigenschaften

$$\frac{\hat{\beta}_{ij} - \beta_{ij}}{\sqrt{\sigma_{jj} a_{ii}}} \sim N(0, 1), \quad (n - p - 1) \hat{\sigma}_{jj} \sim \sigma_{jj} \chi^2(n - p - 1).$$

Simultane Konfidenzbereiche zum globalen Signifikanzniveau  $\alpha_0$  erhält man, wenn man für die individuellen Konfidenzintervalle  $\alpha = \alpha_0/m$  wählt, wobei  $m$  die Anzahl der betrachteten Parameter ist.

### 1.3.3 Tests

Interessante Testprobleme lassen sich als allgemeine lineare Hypothese formulieren. Nullhypothese und Alternativhypothese haben die Form

$$H_0: \mathbf{CB} = \mathbf{I} \quad H_1: \mathbf{CB} \neq \mathbf{I},$$

wobei  $\mathbf{C}$  und  $\mathbf{I}$  fest vorgegebene, geeignet dimensionierte Matrizen sind. Für die Testverfahren wird die Normalverteilungsannahme zugrundegelegt.

Beispiele:

- (1) Will man die Hypothese untersuchen, ob die Variable  $s$  einen Einfluß auf die abhängige hat, formuliert man

$$H_0: \beta_{sj} = 0, j = 1, \dots, q \quad H_1: \beta_{sj} \neq 0 \text{ für mindestens ein } s.$$

Dies läßt sich einfach umformulieren in

$$H_0: (0 \dots 0 \ 1 \ 0 \dots 0) \mathbf{B} = (0 \dots 0),$$

wobei in  $(0 \dots 0 \ 1 \ 0 \dots 0)$  die 1 an der  $(s + 1)$ ten Stelle steht.  $\mathbf{C}$  und  $\mathbf{I}$  sind hier  $(1 \times q)$ -Matrizen.

- (2) Die Globalhypothese „kein Einfluß der Kovariablen auf die abhängige Größe“ besitzt die Form

$$H_0: \beta_{sj} = 0, \quad s = 1, \dots, p, \quad j = 1, \dots, q.$$

In Matrixform ist die äquivalente Nullhypothese

$$H_0: (\mathbf{0}_{p \times 1} \mathbf{I}_{p \times p}) \mathbf{B} = \mathbf{0}_{p \times q}$$

wobei  $\mathbf{0}_{p \times q}$  eine generelle  $(p \times q)$ -Matrix ist, die nur Nullen enthält und  $\mathbf{I}_{p \times p}$  die  $(p \times p)$ -Einheitsmatrix bezeichnet.

Teststatistiken gewinnt man durch den Vergleich der Residuen, die man erhält für das Modell mit der Einschränkung durch  $H_0$  und für das Modell ohne diese Einschränkung.

Bezeichne im weiteren

$\tilde{M}$  das durch die Gültigkeit von  $\mathbf{CB} = \mathbf{T}$  eingeschränkte Modell, und

$M$  das generelle lineare Modell ohne weitere Parameterrestriktion.

Insbesondere vergleicht man die Residuenstreumatrizen

$$SSP(M) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T$$

mit den Residuenstreumatrizen des durch  $H_0$  restringierten Modells  $\tilde{M}$

$$SSP(\tilde{M}) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}})^T,$$

wobei  $\tilde{\boldsymbol{\beta}}$  den Kleinste-Quadrateschätzer unter der Nebenbedingung  $\mathbf{CB} = \mathbf{T}$  bezeichnet.

Als Teststatistiken, die auf dem Vergleich von  $SSP(M)$  und  $SSP(\tilde{M})$  beruhen, benutzt man *Wilks Lambda*

$$\Lambda = \frac{|SSP(M)|}{|SSP(\tilde{M})|},$$

bzw. den Likelihood-Quotienten-Test

$$\lambda = -n \log(\Lambda).$$

Wilks  $\Lambda$  besitzt unter  $H_0$  bei normalverteilten abhängigen Variablen die sogenannte Verteilung Wilks-Verteilung  $\Lambda(q, n - p - 1, s)$ . Die Nullhypothese wird abgelehnt, wenn der Wert von  $\Lambda$  kleiner ist als das  $\alpha$ -Quantil der entsprechenden Wilks-Verteilung  $\Lambda_\alpha(q, n - p - 1, s)$ .

### Streuungszerlegung und Motivation für Wilks Lambda

Bezeichne im Folgenden:

$M$  das multivariate Regressionsmodell und

$\tilde{M}$  das multivariate Regressionsmodell mit der zusätzlichen Forderung, daß  $H_0$  gilt, d.h. der Parameterraum ist durch  $\mathbf{CB} = \mathbf{T}$  restringiert.

Durch Minimierung des Kleinste-Quadratriteriums erhält man die geschätzte Parametermatrix  $\hat{\mathbf{B}}$  mit Komponenten  $\hat{\boldsymbol{\beta}}_{(j)}$ , durch Minimieren des Kleinste-Quadratriteriums mit der zusätzlichen Restriktion  $\mathbf{CB} = \mathbf{T}$  erhält man die Parametermatrix  $\tilde{\mathbf{B}}$  mit

Komponenten  $\tilde{\beta}_{(j)}$ , die jetzt die Restriktion erfüllen. Daraus ergeben sich die Sum-of-squares-and-products (Residuenstreumatrizen) für die Anpassung von  $M$  durch

$$SSP(M) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T = (\mathbf{Y} - \mathbf{X}_0 \hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X}_0 \hat{\mathbf{B}})$$

und für die Anpassung von  $\tilde{M}$  durch

$$SSP(\tilde{M}) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}})^T = (\mathbf{Y} - \mathbf{X}_0 \tilde{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X}_0 \tilde{\mathbf{B}}).$$

Man betrachtet nun die Zerlegung

$$SSP(\tilde{M}) = SSP(M) + SSP(\tilde{M}|M),$$

wobei sich  $SSP(\tilde{M}|M)$  explizit bestimmen läßt durch

$$SSP(\tilde{M}|M) = (\mathbf{C}\hat{\mathbf{B}} - \boldsymbol{\Gamma})^T (\mathbf{C}(\mathbf{X}^T \mathbf{X})\mathbf{C}^T) (\mathbf{C}\hat{\mathbf{B}} - \boldsymbol{\Gamma}).$$

Die Streumatrizen lassen sich als Diskrepanz zwischen Daten und angepaßtem Modell auffassen.  $\tilde{M}$  als das restriktivere Modell besitzt eine schlechtere Anpassung und damit größere Datenabweichung als das generellere Modell  $M$ .  $SSP(\tilde{M}|M)$  ist somit die zusätzliche Abweichung zu den Daten, wenn  $\tilde{M}$  statt  $M$  gefittet wird.

Für den Spezialfall  $q = 1$  und der Hypothese  $\beta_1 = \dots = \beta_p = 0$  entspricht das mit  $\hat{\mathbf{y}}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$  der üblichen Zerlegung

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

Darin wird die Residualstreuung des Modells ohne Kovariablen zerlegt in die Residualstreuung des linearen Modells plus die zusätzliche hinzukommende Streuung wenn das Modell ohne Kovariablen gefittet wird. Der letzte Term wird häufig als die durch das Regressionsmodell erklärte Streuung bezeichnet. Dabei betrachtet man den Übergang von  $\tilde{M}$  zu  $M$ . Wieviel an Variabilität wird durch das Regressionsmodell  $M$  erklärt gegenüber dem Modell ohne Kovariablen  $\tilde{M}$ .

Die Gültigkeit des generelleren Modells vorausgesetzt, folgt  $SSP(M)$  einer Wishart-Verteilung

$$SSP(M) \sim W_q(\boldsymbol{\Sigma}, n - p - 1).$$

Gilt darüberhinaus die Nullhypothese  $H_0: \mathbf{C}\mathbf{B} = \boldsymbol{\Gamma}$ , gilt weiterhin

$$SSP(\tilde{M}|M) \sim W_q(rg(\mathbf{C}))$$

und  $SSP(M)$  und  $SSP(\tilde{M}|M)$  sind unabhängig.

Man erhält somit unter  $H_0$  die unabhängige Zerlegung

$$\begin{array}{rcl}
 SSP(\tilde{M}) & = & SSP(M) \quad + \quad SSP(\tilde{M}|M). \\
 W_q(\boldsymbol{\Sigma}, n-p-1+rg(\mathbf{C})) & & W_q(\boldsymbol{\Sigma}, n-p-M) \quad W_q(\boldsymbol{\Sigma}, rg(\mathbf{C})) \\
 q=1: \sigma^2 \chi^2(n-p-1+rg(\mathbf{C})) & & \sigma^2 \chi^2(n-p-1) \quad \sigma^2 \chi^2(rg(\mathbf{C}))
 \end{array}$$

Mit der Teststatistik

$$\Lambda = \frac{|SSP(M)|}{|SSP(\tilde{M})|} = \frac{|SSP(M)|}{|SSP(M) + SSP(\tilde{M}|M)|}$$

wird die Relevanz von  $SSP(\tilde{M}|M)$  untersucht. Wenn  $H_0$  wahr ist und damit  $\tilde{M}$  eine sinnvolle Reduktion ist, sollte  $SSP(\tilde{M}|M)$  wenig beitragen, andernfalls einen großen Beitrag liefern, so daß  $H_0$  abzulehnen ist, wenn  $\Lambda$  kleine Werte annimmt.

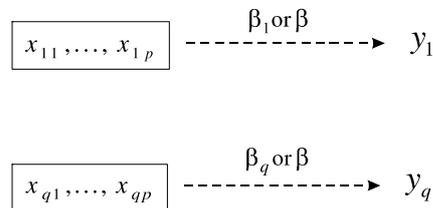


ABBILDUNG 1.2: Marginales Modell

## 1.4 Vorteile des multivariaten linearen Modells

Das multivariate Modell modelliert einen linearen Einfluß der Prädiktoren auf den gesamten Vektor  $\mathbf{y}_i$  und berücksichtigt dabei, daß die Komponenten von  $\mathbf{y}_i$  korreliert sind. Die Korrelation wird insbesondere in den Teststatistiken benutzt. Die Schätzung des multivariaten Modells nach dem Kleinste-Quadrate-Prinzip liefert mit  $\hat{\boldsymbol{\beta}}$  bzw.  $\hat{\mathbf{B}}$  dieselben Schätzungen wie die separate Schätzung der  $q$  multiplen Modelle

$$\mathbf{y}_{(j)} = \mathbf{X}_0 \boldsymbol{\beta}_{(j)} + \boldsymbol{\varepsilon}_{(j)}. \quad (1.5)$$

Was die Schätzung selbst betrifft, könnte man diese multiplen Modelle separat schätzen. Man erhält dann für das Modell (1.5) den Schätzer  $\hat{\boldsymbol{\beta}}_{(j)}$  mit der Kovarianz des Schätzers  $cov(\hat{\boldsymbol{\beta}}_{(j)}) = \sigma_{jj}(\mathbf{X}^T \mathbf{X})^{-1}$ , wobei  $\sigma_{jj}$  aus den Residuen des Modells geschätzt wird. Was erst durch die multivariate Betrachtung verfügbar wird, ist die Schätzung der Kovarianz von  $\hat{\boldsymbol{\beta}}_{(i)}$  und  $\hat{\boldsymbol{\beta}}_{(j)}$ , die durch

$$cov(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(j)}) = \sigma_{ij}(\mathbf{X}_0^T \mathbf{X}_0)^{-1} \quad (1.6)$$

bestimmt ist. Tests für Nullhypothesen der Art

$$H_0: \beta_{(i)} = \beta_{(j)} \quad (1.7)$$

benötigen allerdings die Kenntnis der Schätzgenauigkeit der gemeinsamen Schätzung von  $\hat{\beta}_{(i)}$  und  $\hat{\beta}_{(j)}$ , die nur unter Berücksichtigung der Kovarianz beider Schätzungen angebar ist.

## 1.5 Generellere Modelle

### 1.5.1 Generelles lineares Modell

Das klassische lineare Regressionsmodell geht davon aus, daß ein Vektor von Kovariablen  $\mathbf{x}_i^T = (1, x_{i1}, \dots, x_{ip})$  den Erwartungswert der Komponenten eines abhängigen Vektors bestimmt durch

$$\mu_{ij} = E(y_{ij} | \mathbf{x}_i) = \mathbf{x}_i^T \beta_{(j)}. \quad (1.8)$$

Es wird also der Effekt *eines* Vektors von Einflußgrößen auf  $\mathbf{y}$  spezifiziert, der auf alle Komponenten von  $\mathbf{y}$  wirkt. Das Modell ist allerdings problematisch, wenn komplexere Zusammenhänge zu modellieren sind. Stellen  $y_{i1}, \dots, y_{iq}$  beispielsweise Messungen an aufeinanderfolgenden Tagen dar, wobei ein begleitender Prozeß an Kovariablen  $\mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}$  gemessen wird, ist das Modell (1.8) nicht mehr adäquat. Die Zusammenfassung aller Beobachtungen zu einem großen Vektor  $\mathbf{x}_i^T = (\mathbf{x}_{i1}^T, \dots, \mathbf{x}_{iq}^T)$  wäre denkbar, würde aber in Modell (1.8) implizieren, daß auch zukünftige Kovariablen  $\mathbf{x}_{i,j+1}, \dots, \mathbf{x}_{iq}$  den Erwartungswert von  $y_{ij}$  bestimmen. Häufig ist es daher angebracht, nur den Effekt der Kovariablen  $\mathbf{x}_{ij}$  auf den Erwartungswert  $y_{ij}$  zu modellieren. Man betrachtet dann das *marginale Modell*

$$\mu_{ij} = E(y_{ij} | \mathbf{x}_i) = \mathbf{x}_{ij}^T \beta_{(j)},$$

das den Effekt der Kovariablen  $\mathbf{x}_{ij}$  auf  $y_{ij}$  modelliert.

In einem generellen marginalen Modell ist die *Strukturkomponente* bestimmt durch

$$\boldsymbol{\mu}_i = E(\mathbf{y}_i | \mathbf{X}_i) = \mathbf{X}_i \boldsymbol{\beta},$$

für eine geeignete Designmatrix  $\mathbf{X}_i$ .

Das Modell ist sehr flexibel, die Strukturkomponenten bereits behandelter Modelle ergeben sich daraus als Spezialfälle. Für das oben spezifizierte Modell ist die Designmatrix bestimmt durch

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1}^T & & 0 \\ & \mathbf{x}_{i2}^T & \\ & & \ddots \\ & 0 & & \mathbf{x}_{iq}^T \end{pmatrix}$$

und  $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_{(1)}^T, \dots, \boldsymbol{\beta}_{(g)}^T)$ . Die Einflußgrößenstruktur des klassischen linearen Modells ergibt sich, wenn  $\mathbf{x}_{ij} = \mathbf{x}_i$ .

In geschlossener Matrixschreibweise erhält man für die Strukturkomponente

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta},$$

wobei  $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)$ ,  $\mathbf{X}^T = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)$ .

## 1.5.2 Robuste Schätzung durch generalisierte Schätzgleichungen

### Bekannte Kovarianzmatrix

In einem ersten Schritt nimmt man an, daß die Kovarianz des Responsevektors bekannt ist. In flexibler Form nimmt man an

$$\mathbf{y}_i | \mathbf{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i).$$

Für bekannte Kovarianzmatrix  $\boldsymbol{\Sigma}_i$  läßt sich das Kleinste-Quadrate-Kriterium erweitern zum Gewichteten-Kleinste-Quadrate-Kriterium. Man schätzt  $\boldsymbol{\beta}$  durch Minimieren von

$$WLS(\boldsymbol{\beta}) = \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$$

In Matrixschreibweise hat  $WLS(\boldsymbol{\beta})$  die übersichtliche Form

$$WLS(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

wobei die Gewichtsmatrix eine Block-Diagonalmatrix ist,  $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\Sigma}_n^{-1} \end{pmatrix}$ ,

kurz  $\boldsymbol{\Sigma}^{-1} = \text{Diag}(\boldsymbol{\Sigma}_1^{-1}, \dots, \boldsymbol{\Sigma}_n^{-1})$ . Es läßt sich einfach zeigen, daß Maximieren von  $WLS(\boldsymbol{\beta})$  zu der Schätzgleichung

$$\mathbf{X}^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \quad (1.9)$$

führt, was bei vollem Rang von  $\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}$  den verallgemeinerten Kleinste-Quadrate-Schätzer

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

liefert. Gleichung (1.9) resultiert auch für die Maximum-Likelihood-Schätzung unter Normalverteilungsannahme,  $\hat{\boldsymbol{\beta}}_v$  ist damit auch der ML-Schätzer (bei bekannter Varianz).

### Unbekannte Kovarianzmatrix

Das Hauptproblem bei Verwendung des gewichteten Kleinste-Quadrate-Schätzers ist, daß die Kovarianz von  $\mathbf{y}_i$  und damit  $\Sigma$  meist nicht bekannt sind. Man geht daher nur von der Strukturannahme aus und betrachtet die allgemeine Schätzgleichung mit vorgegebener Gewichtsmatrix  $\mathbf{W}$

$$\mathbf{X}^T \mathbf{W} (\mathbf{y} - \mathbf{X} \hat{\beta}) = 0. \quad (1.10)$$

Die Gewichtsmatrix  $\mathbf{W}$  in der Schätzgleichung ist dabei nicht notwendigerweise die Inverse der wahren Kovarianz, sondern eine *gewählte* Gewichtsmatrix. Der resultierende Schätzer

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

ist der mit  $\mathbf{W}$  assoziierte gewichtete Kleinste-Quadrate-Schätzer, wobei die Existenz von  $(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$  vorausgesetzt wird. Da  $\mathbf{W}^{-1}$  nicht mehr der wahren Kovarianz entspricht, nennt man  $\mathbf{W}^{-1}$  *Arbeits-Kovarianz-Matrix*. Darin kommt zum Ausdruck, daß man nicht annimmt, daß  $\mathbf{W}^{-1}$  tatsächlich die zugrundeliegende Kovarianzmatrix ist,  $\mathbf{W}$  wird nur zur Schätzung verwandt. Wählt man beispielsweise die Gewichtsmatrix  $\mathbf{W} = \mathbf{I}/\sigma^2$ , erhält man unmittelbar den einfachen K-Q-Schätzer

$$\hat{\beta}_{\mathbf{I}/\sigma^2} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Häufig wird man jedoch versuchen,  $\mathbf{W}$  so zu wählen, daß  $\mathbf{W}^{-1}$  näher an der zugrundeliegenden Varianz liegt.

Auch für fest gewählte Gewichtsmatrix hat der Schätzer gute Eigenschaften. Insbesondere ist  $\hat{\beta}_{\mathbf{W}}$  für fest vorgegebene Matrix  $\mathbf{W}$  erwartungstreu, da wegen  $E(\mathbf{y}) = \mathbf{X}\beta$

$$E(\hat{\beta}_{\mathbf{W}}) = \beta$$

gilt. Insbesondere ist auch der einfache Kleinste-Quadrate-Schätzer erwartungstreu. Allerdings ändert sich die Kovarianz des Schätzers und damit seine Effizienz, er hat i.A. nicht mehr die kleinstmögliche Varianz. Für die Varianz ergibt sich bei festem  $\mathbf{W}$  unmittelbar die Sandwich-Matrix

$$\begin{aligned} \text{cov}(\hat{\beta}_{\mathbf{W}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W} \Sigma \mathbf{W} \mathbf{X}) (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \\ &= \mathbf{V}_{\mathbf{W}} \mathbf{V}_{\Sigma} \mathbf{V}_{\mathbf{W}}, \end{aligned}$$

wobei

$$\mathbf{V}_{\mathbf{W}} = \mathbf{X}^T \mathbf{W} \mathbf{X} \quad \text{und} \quad \mathbf{V}_{\Sigma} = \mathbf{X}^T \mathbf{W} \Sigma \mathbf{W} \mathbf{X}.$$

Während  $\mathbf{V}_{\mathbf{W}}$  bekannt ist, enthält  $\mathbf{V}_{\Sigma}$  die unbekannte Kovarianzmatrix von  $\mathbf{y}$ . Dies ist jedoch weit weniger problematisch als es scheint. Ein robuster Schätzer für die Kovarianz von  $\hat{\beta}_{\mathbf{W}}$  ergibt sich, wenn  $\Sigma$  durch eine konsistente Schätzung  $\hat{\Sigma}$  ersetzt wird. Ein Kandidat ist

$$\hat{\Sigma} = \frac{1}{n-p} \sum_{i=1}^n (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^T$$

---

Zu betonen ist, dass der Schätzer nur die Strukturannahme  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  macht. Es ist auch nichts über die Verteilung vorausgesetzt. Er bietet sich vor allem dann an, wenn Zweifel an der Normalitätsannahme oder der Varianzhomogenität bestehen.

Spezifikation der Kovarianzmatrix