

---

**Forschungszentrum Jülich GmbH**

**Jülich Supercomputing Centre (JSC)**

---

**Statistische Modellierung**  
**Vorlesungsskript**

Wolfgang Meyer

Jülich, 5. Juni 2014



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Grundlagen</b>	<b>1</b>
1.1	Einführung des linearen Modells . . . . .	1
1.2	Lineare Algebra	
	Orthogonalprojektion und verallgemeinerte Inverse . . . . .	9
1.3	Wahrscheinlichkeitsrechnung	
	Mehrdimensionale stetige Verteilungen . . . . .	18
<b>2</b>	<b>Das lineare Modell</b>	<b>25</b>
2.1	Die Methode der kleinsten Quadrate und Schätzbarkeit . . . . .	25
2.2	Streuungszerlegung und kanonische Darstellung . . . . .	45
<b>3</b>	<b>Das klassische lineare Modell</b>	<b>51</b>
3.1	Die mehrdimensionale Normalverteilung . . . . .	51
3.2	$\chi^2$ -, F- und t-Verteilung . . . . .	56
3.3	Die Maximum-Likelihood-Methode und die Verteilung der Schätzer . . . . .	62
<b>4</b>	<b>Konfidenzbereiche</b>	<b>67</b>
4.1	Konfidenzintervalle und Vorhersageintervalle . . . . .	67
4.2	Simultane Konfidenzintervalle und gemeinsame Konfidenzbereiche . . . . .	74
<b>5</b>	<b>Hypothesentests</b>	<b>81</b>
5.1	F-Test für eine testbare Hypothese . . . . .	81
5.2	Tests spezieller Hypothesen . . . . .	90
<b>6</b>	<b>Varianzanalyse</b>	<b>103</b>
6.1	Einfache Varianzanalyse . . . . .	103
6.2	Zweifache Varianzanalyse . . . . .	109



# Kapitel 1

## Einleitung und Grundlagen

### 1.1 Einführung des linearen Modells

Das Ziel wissenschaftlicher Untersuchungen besteht häufig in der Aufdeckung und Analyse von Zusammenhängen. Beispiele hierfür sind

1. das Ohmsche Gesetz  $U = R \cdot I$ ,  $U$  Spannung,  $R$  Widerstand,  $I$  Stromstärke,
2. das Fallgesetz  $s = \frac{1}{2} \cdot g \cdot t^2$ ,  $s$  Weg,  $g$  Erdbeschleunigung,  $t$  Zeit,
3. die Kondensatorentladung  $u(t) = u(0) \cdot e^{-at}$ ,  $u(t)$  Spannung zur Zeit  $t$ ,  $a$  kondensatorspezifischer Parameter,
4. die Michaelis-Menten Gleichung der Enzymkinetik (Umwandlung einer Ausgangssubstanz, das sogenannte Substrat, in eine veränderte Substanz, das sogenannte Produkt)

$$V = \frac{V_{max}S}{K_m + S} ,$$

$S$  Menge des Substrats,  $V$  Reaktionsgeschwindigkeit (Abbaurate des Substrats bzw. Produktionsrate des Produkts),  $V_{max}$  und  $K_M$  reaktionsspezifische Parameter,

5. die Abhängigkeit des Blutdrucks vom Alter,
6. die Festigkeit eines Materials in Abhängigkeit von der Temperatur und dem Druck während der Produktion,
7. der Vergleich der Töchterleistungen von drei Bullen bezüglich der Milchfettmenge.

Die Beispiele 1 bis 4 stellen funktionale Beziehungen der Form  $z = f(x; \beta)$  dar mit einer abhängigen Variablen  $z$ , einer unabhängigen Variablen  $x$  und einem häufig unbekannten Parameter oder Parametervektor  $\beta$ , der gegebenenfalls mittels der Beobachtung von Wertepaaren  $(x, z)$  zu bestimmen ist. Können  $x$  und  $z$  exakt bestimmt werden, so liegt ein deterministisches Modell vor. Häufig ist eine

Größe, die im allgemeinen als abhängige Veränderliche  $z$  gewählt wird, nicht exakt bestimmbar sondern mit einem (zufälligen) Messfehler  $\epsilon$  versehen, das heisst, es wird nicht  $z$  sondern  $Y = z + \epsilon$  bestimmt, so werden aus den deterministischen Modellen 1 bis 4 die folgenden stochastischen Modelle mit der deterministischen Komponente  $f(x, \beta)$  und der stochastischen  $\epsilon$ :

$$1^*. Y = f(x; \beta) + \epsilon = \beta x + \epsilon, \quad Y = I + \epsilon, \quad z = I, \quad x = U, \quad \beta = \frac{1}{R},$$

$$2^*. Y = f(x; \beta) + \epsilon = \beta x^2 + \epsilon, \quad Y = s + \epsilon, \quad z = s, \quad x = t, \quad \beta = \frac{1}{2} \cdot g,$$

$$3^*. Y = f(x; \beta) + \epsilon = \beta_1 \cdot e^{-\beta_2 x} + \epsilon, \quad Y = u(t) + \epsilon, \quad z = u(t), \quad x = t, \quad \beta = (u(0), a)^T$$

$$4^*. Y = f(x; \beta) + \epsilon = \frac{\beta_1 x}{\beta_2 + x} + \epsilon, \quad Y = V + \epsilon, \quad z = V, \quad x = S, \quad \beta = (V_{max}, K_m)^T$$

Der Blutdruck eines Menschen steigt erfahrungsgemäß mit zunehmendem Alter an, es besteht aber kein funktionaler Zusammenhang zwischen Blutdruck und Alter. Durch das Alter wird der Blutdruck nicht eindeutig festgelegt, er wird von vielen weiteren nicht beobachteten oder nicht beobachtbaren Größen sowie von „zufälligen“ Eigenschaften der Untersuchungseinheit beeinflusst. Den Blutdruck einer zufällig ausgewählten Person mit Alter  $x$  (zu einem bestimmten Zeitpunkt) können wir beschreiben durch eine Zufallsgröße  $Z$  deren Erwartungswert eine Funktion  $\mu(x)$  des Alters  $x$  ist. Der gemessene Blutdruck  $Y$  ist dann der von einem Messfehler  $\epsilon_m$  überlagerte tatsächliche Wert:

$$Y = Z + \epsilon_m = \mu(x) + \epsilon_Z + \epsilon_m = \mu(x) + \epsilon$$

Für die unbekannte Funktion  $\mu(x)$  wählen wir eine Modellfunktion  $f(x; \beta)$  mit einem unbekannten Parameter oder Parametervektor  $\beta$ , von der wir annehmen, dass sie  $\mu(x)$  auf dem zu untersuchenden Altersbereich hinreichend approximiert. So erhalten wir zum Beispiel das folgende stochastische Modell für Beispiel 5:

$$5^*. Y = f(x; \beta) + \epsilon = \beta_1 + \beta_2 x + \epsilon, \quad Y \text{ ist der gemessene Blutdruck, } x \text{ das Alter.}$$

Entsprechend erhalten wir für Beispiel 6, die Festigkeit eines Materials in Abhängigkeit von Temperatur und Druck, das stochastische Modell

$$6^*. Y = f(x_1, x_2; \beta) + \epsilon = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_1 x_2 + \epsilon,$$

$Y$  ist die gemessene Festigkeit,  $x_1$  die Temperatur,  $x_2$  der Druck.

Das Beispiel zur Milchfettproduktion (7) unterscheidet sich von dem Blutdruckbeispiel (5) durch das Merkmal bzw. den Faktor, der die Untersuchungseinheiten charakterisiert und dessen Einfluss auf die abhängige Größe untersucht werden soll. Im Falle des Blutdrucks ist dieses das Alter, eine quantitative Größe, und bei der Milchfettproduktion (7) ist es der Bulle, von dem die Kühe abstammen, eine qualitative Größe mit drei Merkmalsausprägungen. Die Fettmenge, die von einer Tochter des  $j$ -ten Bullen ( $j = 1, 2, 3$ ) produziert wird, beschreiben wir durch eine Zufallsgröße  $Z$  mit einem Erwartungswert  $\mu_j$ . Die gemessene Fettmenge  $Y$  ist der von einem Messfehler  $\epsilon_m$  überlagerte tatsächliche Wert:

$$Y = Z + \epsilon_m = \mu_j + \epsilon_Z + \epsilon_m = \mu_j + \epsilon.$$

Durch die Einführung sogenannter Dummy-Variablen können wir dieses stochastische Modell formulieren wie die bisher betrachteten. Es sei  $x_j = 1$  oder  $0$  ( $j = 1, 2, 3$ ), je nachdem ob die untersuchte Kuh von dem  $j$ -ten Bullen abstammt oder nicht, dann gilt:

$$7^*. Y = f(x_1, x_2, x_3; \beta) + \epsilon = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon,$$

$Y$  ist die gemessene MilCHFettmenge,  $\beta_j = \mu_j$  die (unbekannte) mittlere MilCHFettmenge einer Tochter des  $j$ -ten Bullen.

Von einem linearen Modell spricht man, wenn die deterministische Komponente des Modells  $f(x_1, \dots, x_l; \beta)$  in dem Parameter bzw. in den Parametern linear ist und der Erwartungswert der stochastischen Komponente  $\epsilon$ , die als Fehler bezeichnet wird, die Voraussetzung  $E[\epsilon] = 0$  erfüllt. Lineare Modelle in diesem Sinne sind die Modelle  $1^*$ ,  $2^*$ ,  $5^*$ ,  $6^*$  und  $7^*$ , sofern  $E[\epsilon] = 0$  gilt. Die deterministischen Modellfunktionen der Modelle  $1^*$ ,  $5^*$  und  $7^*$  sind außerdem lineare Funktionen, das heißt, sie sind linear in  $x_1, \dots, x_l$ . Die Modellfunktion  $f$  in  $2^*$  ist quadratisch in  $x$  und die in  $6^*$  ist ein Polynom zweiten Grades in  $x_1$  und  $x_2$ , es liegen also (falls  $E[\epsilon] = 0$ ) lineare Modelle mit nichtlinearen Modellfunktionen  $f$  vor. Die Modellfunktionen  $f$  in  $3^*$  und  $4^*$  sind nichtlinear in den Parametern und in der unabhängigen Veränderlichen.

Logarithmieren wir im Beispiel 3 beide Seiten der Gleichung und betrachten den funktionalen Zusammenhang von  $\tilde{u}(t) = \log u(t)$  und  $t$ , so ergibt sich  $\tilde{u}(t) = \tilde{u}(0) - at$ . Ausgehend von dieser Gleichung erhalten wir ein zu  $3^*$  alternatives stochastisches Modell

$$3^{**}. \tilde{Y} = \tilde{f}(x; \tilde{\beta}) + \tilde{\epsilon} = \tilde{\beta}_1 + \tilde{\beta}_2 x + \tilde{\epsilon}, \quad \tilde{Y} = \log u(t) + \tilde{\epsilon}, \quad x = t, \quad \tilde{\beta} = (\log u(0), -a)^T$$

Die deterministische Komponente dieses Modells ist linear in den Parametern und in der unabhängigen Veränderlichen. In wieweit sich eines dieser beiden Modelle als Grundlage für eine statistische Analyse eignet, hängt ab von den Eigenschaften der Fehlerterme  $\epsilon$  bzw.  $\tilde{\epsilon}$ .

Statistische Aussagen über die unbekannten Parameter eines dieser stochastischen Modelle sind nur möglich, wenn hinreichend viele Beobachtungen zu verschiedenen Faktorkombinationen vorliegen und die Fehler in den Beobachtungen weitere Voraussetzungen erfüllen. Es stehen also  $n$  (i.a.  $n \geq k$ ) Messungen mit entsprechenden Faktorkombinationen für das lineare Modell in seiner allgemeinsten Form

$$\begin{aligned} Y &= f(\tilde{x}_1, \dots, \tilde{x}_l; \beta_1, \dots, \beta_k) + \epsilon \\ &= \beta_1 q_1(\tilde{x}_1, \dots, \tilde{x}_l) + \dots + \beta_k q_k(\tilde{x}_1, \dots, \tilde{x}_l) + \epsilon, \quad E[\epsilon] = 0 \end{aligned} \tag{1.1.1}$$

zur Verfügung:

$$\begin{aligned} Y_i &= f(\tilde{x}_{i1}, \dots, \tilde{x}_{il}; \beta_1, \dots, \beta_k) + \epsilon_i \\ &= \beta_1 q_1(\tilde{x}_{i1}, \dots, \tilde{x}_{il}) + \dots + \beta_k q_k(\tilde{x}_{i1}, \dots, \tilde{x}_{il}) + \epsilon_i, \quad E[\epsilon_i] = 0 \\ i &= 1, \dots, n \end{aligned} \tag{1.1.2}$$

Setzen wir  $\vec{Y} = (Y_1, \dots, Y_n)^T$ ,  $\vec{\beta} = (\beta_1, \dots, \beta_k)^T$ ,  $\vec{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ ,  $E[\vec{\epsilon}] = (E[\epsilon_1], \dots, E[\epsilon_n])^T$  und  $X = (x_{ij})_{i=1, \dots, n, j=1, \dots, k}$  mit  $x_{ij} = q_j(\tilde{x}_{i1}, \dots, \tilde{x}_{il})$ , dann können die Modellgleichungen (1.1.2) im Matrizenkalkül formuliert werden:

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad (1.1.3)$$

Üblicherweise wird angenommen, dass alle Fehler die gleiche Varianz haben („Homoskedastie“) und paarweise unkorreliert sind. Unter diesen Voraussetzungen können Aussagen über die Genauigkeit von Parameterschätzungen gemacht werden sowie bei unabhängigen und identisch verteilten Fehlern und geeigneter Wahl der Faktorwerte asymptotische Tests durchgeführt und asymptotische Konfidenzbereiche bestimmt werden.

Setzt man zusätzlich voraus, dass die Fehler normalverteilt sind (Klassisches Lineares Modell), dann sind auch die Parameterschätzungen normalverteilt und es können exakte Tests durchgeführt und exakte Konfidenzintervalle bestimmt werden.

Diese Voraussetzungen können abgeschwächt werden, so reicht es unter Umständen aus vorauszusetzen, dass die Kovarianzmatrix der Fehler bis auf eine Konstante bekannt ist.

### Definition 1.1.1. Lineares Modell

1. Das (allgemeine) lineare Modell (mit festen Effekten) ist charakterisiert durch die Gleichungen

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = (Cov(\epsilon_i, \epsilon_j))_{i=1, \dots, n, j=1, \dots, n} = \sigma^2 I, \quad \vec{\beta} \in \mathbb{R}^k. \quad (1.1.4)$$

Dabei ist  $\vec{Y} = (Y_1, \dots, Y_n)^T$  der Zufallsvektor der beobachteten Größen (abhängige Variable),  $\vec{\beta} = (\beta_1, \dots, \beta_k)^T$  der Vektor der unbekannten konstanten Parameter,  $X = (x_{ij})_{i=1, \dots, n, j=1, \dots, k}$  die Matrix der konstanten Werte der Einflussfaktoren (kontrollierte oder unabhängige Variablen), auch Designmatrix genannt,  $\vec{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  der Zufallsvektor der Fehler (Störgrößen),  $\sigma^2 > 0$  die Varianz der Fehlergrößen,  $I$  die  $(n \times n)$ -Einheitsmatrix.

2. Man sagt, eine nicht konstante kontrollierte Größe gehört zu einem qualitativen Faktor, wenn die zugehörige Spalte nur aus Nullen und Einsen besteht, und spricht andernfalls von einem quantitativen Faktor.

Man spricht beim linearen Modell von

- Regressionsanalyse, falls alle nicht konstanten kontrollierten Größen zu quantitativen Faktoren gehören,
- Varianzanalyse (englisch: analysis of variance, Abkürzung: ANOVA), falls alle nicht konstanten kontrollierten Größen zu qualitativen Faktoren gehören,
- Kovarianzanalyse, falls mindestens eine kontrollierte Größe zu einem quantitativen Faktor und mindestens eine zu einem qualitativen Faktor gehört.



**Bemerkung 1.**

Für das durch Gleichung (1.1.4) gegebene lineare Modell gilt

$$E[\vec{Y}] = X\vec{\beta} = \sum_{j=1}^k \vec{x}_j \beta_j, \quad \Sigma_{\vec{Y}} = \sigma^2 I, \quad (1.1.5)$$

wobei  $\vec{x}_1, \dots, \vec{x}_k$  die Spaltenvektoren von  $X$  sind. Die Menge der aufgrund des Modells möglichen Erwartungswerte des Beobachtungsvektors  $\vec{Y}$  bildet den Spaltenraum von  $X$ . Der Parametervektor  $\vec{\beta}$  wird durch den Erwartungswert von  $\vec{Y}$  genau dann eindeutig festgelegt, wenn die Spaltenvektoren von  $X$  linear unabhängig sind, d. h. wenn die Matrix  $X$  vollen Spaltenrang hat,  $\text{rg}(X) = k$ .

**Bemerkung 2. Alternative parameterfreie Formulierung des linearen Modells**

Es sei  $\mathcal{U}$  der Spaltenraum der Designmatrix  $X$ :

$$\mathcal{U} = \{\vec{\mu} \mid \vec{\mu} = X\vec{\beta}, \vec{\beta} \in \mathbb{R}^k\}. \quad (1.1.6)$$

Das lineare Modell kann dann wie folgt parameterfrei formuliert werden:

$$\vec{Y} = \vec{\mu} + \vec{\epsilon}, \quad E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\mu} \in \mathcal{U} \quad (1.1.7)$$

Dabei ist  $\vec{Y} = (Y_1, \dots, Y_n)^T$  der Zufallsvektor der beobachteten Größen (abhängige Variable),  $\vec{\mu} = (\mu_1, \dots, \mu_n)^T$  der Vektor der unbekannten Erwartungswerte der beobachteten Größen,  $\vec{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  der Zufallsvektor der Fehler (Störgrößen),  $\sigma^2 > 0$  die Varianz der Fehlergrößen,  $I$  die  $(n \times n)$ -Einheitsmatrix.

Stimmen die Spaltenräume der  $(n \times k)$ -Matrix  $X$  und der  $(n \times k_*)$ -Matrix  $X_*$  überein, dann beschreiben die linearen Modelle

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\beta} \in \mathbb{R}^k \quad (1.1.8)$$

sowie

$$\vec{Y} = X_*\vec{\beta}_* + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\beta}_* \in \mathbb{R}^{k_*} \quad (1.1.9)$$

dasselbe parameterfreie lineare Modell. Sie stellen unterschiedliche Parametrisierungen des durch Gleichung (1.1.7) gegebenen linearen Modells dar. Die Änderung der Parametrisierung wird als Re-parametrisierung bezeichnet.

**Beispiel 1.1.1. Einfache lineare Regression**

Für die zugrundeliegende Modellfunktion

$$f(x; \beta_0, \beta_1) = \beta_0 + \beta_1 x \quad (1.1.10)$$

liegen Beobachtungen  $Y_1, \dots, Y_n$  der Funktionswerte für die  $x$ -Werte  $x_1, \dots, x_n$  vor. Die Modellfunktion ist eine einfache lineare Regressionsfunktion mit dem Achsenabschnitt bzw. der Regressionskonstanten  $\beta_0$  und dem Regressionskoeffizienten  $\beta_1$ . Das durch die Gleichungen

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i, \quad E[\epsilon_i] = 0, \quad \text{Var}(\epsilon_i) = \sigma^2 > 0 \quad (i = 1, \dots, n) \\ \text{Kov}(\epsilon_i, \epsilon_j) &= 0 \quad (1 \leq i < j \leq n) \end{aligned} \quad (1.1.11)$$

gegebene lineare Modell lautet in Matrizenschreibweise:

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\epsilon} = \sigma^2 I, \quad (1.1.12)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \quad (1.1.13)$$

Die Modellfunktion heißt *Regressionsgerade* oder *Ausgleichsgerade*,  $\beta_0$  ist der *Achsenabschnitt* oder die *Regressionskonstante* (englisch: *intercept*) und  $\beta_1$  der *Regressionskoeffizient*. Die *unabhängige Variable*  $x$  wird auch als *Regressor* oder *Einflussgröße* und die *abhängige Variable*  $Y$  als *Regressand* oder *Zielgröße* bezeichnet.

### Beispiel 1.1.2. Multiple oder mehrfache lineare Regression

Für die zugrundeliegende Modellfunktion

$$f(x_1, \dots, x_k; \beta_0, \dots, \beta_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (1.1.14)$$

liegen Beobachtungen  $Y_1, \dots, Y_n$  der Funktionswerte für die  $(x_1, \dots, x_k)$ -Werte  $(x_{1,1}, \dots, x_{1,k}), \dots, (x_{n,1}, \dots, x_{n,k})$  vor. Die Modellfunktion ist eine *multiple lineare Regressionsfunktion* mit dem *Achsenabschnitt* bzw. der *Regressionskonstanten*  $\beta_0$  und den (*partiellen*) *Regressionskoeffizienten*  $\beta_1, \dots, \beta_k$ . Das durch die Gleichungen

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k} + \epsilon_i, \quad E[\epsilon_i] = 0, \\ \text{Var}(\epsilon_i) &= \sigma^2 > 0 \quad (i = 1, \dots, n), \quad \text{Kov}(\epsilon_i, \epsilon_j) = 0 \quad (1 \leq i < j \leq n) \end{aligned} \quad (1.1.15)$$

gegebene lineare Modell lautet in Matrizenschreibweise:

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\epsilon} = \sigma^2 I, \quad (1.1.16)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,k} \\ 1 & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k} \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \quad (1.1.17)$$

Die Modellfunktion heißt *lineare Regressionsfunktion* oder *Ausgleichsfunktion*,  $\beta_0$  ist der *Achsenabschnitt* oder die *Regressionskonstante* (englisch: *intercept*), und  $\beta_1, \dots, \beta_k$  sind die (*multiplen* oder *partiellen*) *Regressionskoeffizienten*. Die *unabhängigen Variablen*  $x_1, \dots, x_k$  werden auch als *Regresoren* oder *Einflussgrößen* und die *abhängige Variable*  $Y$  als *Regressand* oder *Zielgröße* bezeichnet.

**Beispiel 1.1.3. Polynomiale Regression**

Für die zugrundeliegende Modellfunktion

$$f(x; \beta_0, \dots, \beta_k) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k \quad (1.1.18)$$

liegen Beobachtungen  $Y_1, \dots, Y_n$  der Funktionswerte für die  $x$ -Werte  $x_1, \dots, x_n$  vor. Das durch die Gleichungen

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \epsilon_i, \quad E[\epsilon_i] = 0, \\ \text{Var}(\epsilon_i) &= \sigma^2 > 0 \quad (i = 1, \dots, n), \quad \text{Kov}(\epsilon_i, \epsilon_j) = 0 \quad (1 \leq i < j \leq n) \end{aligned} \quad (1.1.19)$$

gegebene lineare Modell lautet in Matrixschreibweise:

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad (1.1.20)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \quad (1.1.21)$$

**Beispiel 1.1.4. Einfache Varianzanalyse**

Um  $k$  experimentelle Bedingungen (z.B. Behandlungen, Töchterleistungen verschiedener Bullen (Beispiel auf Seite 2)) zu vergleichen, wird die  $i$ -te Behandlung auf  $n_i > 0$  verschiedene Versuchseinheiten angewandt. Dabei wird vorausgesetzt, dass sich die unterschiedlichen Bedingungen nur auf die Erwartungswerte, nicht aber auf die Variabilität der Messungen auswirken. Dieser Versuchsplan führt auf das  $k$ -Stichprobenproblem, dem einfachsten Problem der Varianzanalyse.

Es liegt ein Einflussfaktor (Behandlungsart, Abstammung (Vaterbulle)) mit  $k$  Ausprägungen, Stufen (englisch: levels) vor, daher spricht man von einer einfachen, einfaktoriellen Varianzanalyse.

Bezeichnet  $Y_{ij}$  ( $j = 1, \dots, n_i, i = 1, \dots, k$ ) den  $j$ -ten Messwert unter der  $i$ -ten Bedingung und  $\mu_i$  ( $i = 1, \dots, k$ ) den Erwartungswert bei Vorliegen der  $i$ -ten Bedingung, dann kann das zugrundeliegende lineare Modell wie folgt formuliert werden:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad E[\epsilon_{ij}] = 0, \quad \text{Var}(\epsilon_{ij}) = \sigma^2 > 0, \quad (1.1.22)$$

$$\text{Kov}(\epsilon_{ij}, \epsilon_{i'j'}) = 0 \quad (i, j) \neq (i', j') \quad (j = 1, \dots, n_i, j' = 1, \dots, n_{i'}, i, i' = 1, \dots, k)$$

Mit  $\vec{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$  und  $\vec{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$  für  $i = 1, \dots, k$  ergibt sich in Matrizenform:

$$\vec{Y} = \tilde{X}\vec{\gamma} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad (1.1.23)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_k \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \vec{1}_{n_1} & \vec{0}_{n_1} & \dots & \vec{0}_{n_1} \\ \vec{0}_{n_2} & \vec{1}_{n_2} & \dots & \vec{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{0}_{n_k} & \vec{0}_{n_k} & \dots & \vec{1}_{n_k} \end{bmatrix}, \quad \vec{\gamma} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{bmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \\ \vdots \\ \vec{\epsilon}_k \end{bmatrix} \quad (1.1.24)$$

Die Designmatrix  $\tilde{X}$  hat vollen Spaltenrang.

Häufig wird eine alternative Parametrisierung verwendet, bei der die Erwartungswerte in eine gemeinsame Komponente  $\mu$  und Abweichungen  $\alpha_i$  zerlegt werden, d.h.  $\mu_i = \mu + \alpha_i$  für  $i = 1, \dots, k$ .

Die Modellgleichungen lauten dann

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + \epsilon_{ij}, \quad E[\epsilon_{ij}] = 0, \quad \text{Var}(\epsilon_{ij}) = \sigma^2 > 0, \\ \text{Kov}(\epsilon_{ij}, \epsilon_{i'j'}) &= 0 \quad (i, j) \neq (i', j') \quad (j = 1, \dots, n_i, j' = 1, \dots, n_{i'}, i, i' = 1, \dots, k) \end{aligned} \quad (1.1.25)$$

und in Matrizenform

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad (1.1.26)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_k \end{bmatrix}, \quad X = \begin{bmatrix} \vec{1}_{n_1} & \vec{1}_{n_1} & \vec{0}_{n_1} & \dots & \vec{0}_{n_1} \\ \vec{1}_{n_2} & \vec{0}_{n_2} & \vec{1}_{n_2} & \dots & \vec{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{1}_{n_k} & \vec{0}_{n_k} & \vec{0}_{n_k} & \dots & \vec{1}_{n_k} \end{bmatrix}, \quad \vec{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{bmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \\ \vdots \\ \vec{\epsilon}_k \end{bmatrix} \quad (1.1.27)$$

Die Designmatrix  $X$  hat nicht vollen Spaltenrang. Die Zerlegung der Erwartungswerte in dieser Form ist nicht eindeutig, so dass zusätzliche Nebenbedingungen an die Parameter  $\mu, \alpha_1, \dots, \alpha_k$  gestellt werden müssen. Man fordert in diesem Fall gemeinhin  $\sum_{i=1}^k \alpha_i = 0$  oder  $\sum_{i=1}^k n_i \alpha_i = 0$ . Im ersten Fall ist  $\mu = \frac{1}{k} \sum_{i=1}^k \mu_i$ , der über die  $k$  Bedingungen gemittelte Erwartungswert, und im zweiten  $\mu = \frac{1}{n} \sum_{i=1}^k n_i \mu_i = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} E[Y_{ij}]$ , der über alle  $n = \sum_{i=1}^k n_i$  Beobachtungen gemittelte Erwartungswert. Die Nebenbedingungen, durch die die Parameter eindeutig festgelegt werden, bezeichnet man als Reparametrisierungsbedingungen.

In diesem Beispiel wird nicht ersichtlich, weshalb diese alternative Parametrisierung gewählt wird. Erst in komplexeren Beispielen zeigt sich ihre Nützlichkeit.

### Bemerkung 1.

Enthält eine (i.a. die erste) Spalte der Designmatrix eines linearen Modells nur Einsen, so spricht man von einem Modell mit Achsenabschnitt oder Intercept-Modell und bezeichnet den zugehörigen Parameter als Achsenabschnitt oder Intercept. Enthält der Spaltenraum der Designmatrix eines linearen Modells den Vektor  $\vec{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$ , dann ist eine Parametrisierung mit Intercept-Parameter möglich.

## 1.2 Lineare Algebra

### Orthogonalprojektion und verallgemeinerte Inverse

#### Definition 1.2.1.

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum und  $g : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  eine Abbildung.

1.  $g$  ist eine Bilinearform genau dann, wenn für alle  $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in \mathcal{V}$  und  $c_1, c_2 \in \mathbb{R}$  gilt:

$$\begin{aligned} g(c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2, \mathbf{w}) &= c_1 g(\mathbf{v}_1, \mathbf{w}) + c_2 g(\mathbf{v}_2, \mathbf{w}) \\ g(\mathbf{v}, c_1 \mathbf{w}_1 + c_2 \mathbf{w}_2) &= c_1 g(\mathbf{v}, \mathbf{w}_1) + c_2 g(\mathbf{v}, \mathbf{w}_2) \end{aligned} \quad (1.2.1)$$

2. Eine Bilinearform  $g$  ist positiv definit genau dann, wenn

$$g(\mathbf{v}, \mathbf{v}) \geq 0, \quad g(\mathbf{v}, \mathbf{v}) = 0 \Leftrightarrow \mathbf{v} = \mathbf{o} \quad \forall \mathbf{v} \in \mathcal{V} \quad (1.2.2)$$

3. Eine Bilinearform  $g$  ist symmetrisch genau dann, wenn

$$g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v}) \quad \forall \mathbf{v}, \mathbf{w} \in \mathcal{V} \quad (1.2.3)$$

4. Ein Skalarprodukt ist eine positiv-definite symmetrische Bilinearform.

Schreibweise:

$$g(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle \quad (1.2.4)$$

#### Satz 1.2.1.

1. Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum und  $g$  eine Bilinearform auf  $\mathcal{V}$ . Dann gibt es zu jeder Basis  $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$  von  $\mathcal{V}$  eine quadratische Matrix  $A$  der Ordnung  $n$  mit

$$g(\mathbf{v}, \mathbf{w}) = \vec{x}^T A \vec{y} \quad (1.2.5)$$

wobei  $\vec{x} = (x_1, \dots, x_n)^T, \vec{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$  die Koordinatenvektoren der Vektoren  $\mathbf{v}, \mathbf{w} \in \mathcal{V}$  bezüglich der Basis sind. Es ist  $A = (g(\mathbf{a}_i, \mathbf{a}_j))_{i,j=1,\dots,n}$ .

2. Es sei  $A$  die Matrix einer Bilinearform  $g$  auf einem endlich-dimensionalen reellen Vektorraum  $\mathcal{V}$ . Dann gilt

- $g$  ist genau dann positiv-definit, wenn  $A$  positiv definit ist.
- $g$  ist genau dann symmetrisch, wenn  $A$  symmetrisch ist.

**Definition 1.2.2.**

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt.

1. Zwei Vektoren  $\mathbf{v}, \mathbf{w} \in \mathcal{V}$  heißen orthogonal genau dann, wenn  $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ .

Schreibweise:  $\mathbf{v} \perp \mathbf{w}$

2. Es sei  $S$  eine Teilmenge von  $\mathcal{V}$ , so bezeichnet  $S^\perp$  die Menge aller zu  $S$  orthogonalen Vektoren:

$$S^\perp = \{\mathbf{v} \in \mathcal{V} \mid \langle \mathbf{v}, \mathbf{s} \rangle = 0 \ \forall \mathbf{s} \in S\} \quad (1.2.6)$$

3.  $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$  heißt Länge oder Norm von  $\mathbf{v} \in \mathcal{V}$ .

4. Ein Vektor  $\mathbf{v} \in \mathcal{V}$  heißt normiert genau dann, wenn  $\|\mathbf{v}\| = 1$ .

5. Zwei orthogonale normierte Vektoren heißen orthonormal.

6. Eine Menge von Vektoren, die paarweise orthogonal (orthonormal) sind, heißt orthogonal (orthonormal) oder Orthogonalsystem (Orthonormalsystem). Eine Basis aus orthogonalen (orthonormalen) Vektoren heißt Orthogonalbasis (Orthonormalbasis).

**Satz 1.2.2.**

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt.

1.  $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2 \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}$

2. Satz von Pythagoras:  $\mathbf{v} \perp \mathbf{w} \Leftrightarrow \|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}$

3. Gram-Schmidtsches Orthogonalisierungsverfahren:

Es sei  $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$  eine Basis von  $\mathcal{V}$ , dann bilden die wie folgt definierten Vektoren  $\mathbf{b}_1, \dots, \mathbf{b}_n$  eine Orthogonalbasis von  $\mathcal{V}$ :

$$\mathbf{b}_1 = \mathbf{a}_1, \quad \mathbf{b}_j = \mathbf{a}_j - \sum_{i=1}^{j-1} \frac{\langle \mathbf{a}_j, \mathbf{b}_i \rangle}{\langle \mathbf{b}_i, \mathbf{b}_i \rangle} \cdot \mathbf{b}_i \quad (2 \leq j \leq n) \quad (1.2.7)$$

4. Sind  $\vec{x} = (x_1, \dots, x_n)^T, \vec{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$  die Koordinatenvektoren der Vektoren  $\mathbf{v}, \mathbf{w} \in \mathcal{V}$  bezüglich einer Orthonormalbasis von  $\mathcal{V}$ , dann gilt:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \vec{x}^T \vec{y} = \sum_{i=1}^n x_i y_i \quad (1.2.8)$$

$\vec{x}^T \vec{y}$  ist das kanonische Skalarprodukt der Vektoren  $\vec{x}$  und  $\vec{y}$ .

5. Es sei  $S$  eine Teilmenge von  $\mathcal{V}$  und  $\mathcal{L}(S)$  der von  $S$  in  $\mathcal{V}$  aufgespannte bzw. erzeugte Vektorraum, dann gilt:

$$S^\perp = \mathcal{L}(S)^\perp \quad (1.2.9)$$

**Satz 1.2.3. Projektionssatz**

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt,  $\mathcal{U}$  ein Untervektorraum von  $\mathcal{V}$  und  $\mathbf{v} \in \mathcal{V}$ .

Es gibt genau einen Vektor  $\mathbf{u}_0 \in \mathcal{U}$ , sodass  $\mathbf{v} - \mathbf{u}_0$  orthogonal zu  $\mathcal{U}$  ist:

$$(\mathbf{v} - \mathbf{u}_0) \perp \mathbf{u} \quad \forall \mathbf{u} \in \mathcal{U}. \quad (1.2.10)$$

Das heißt, es existiert eine eindeutige Darstellung von  $\mathbf{v}$  der Form

$$\mathbf{v} = \mathbf{u} + \mathbf{w}, \quad \mathbf{u} \in \mathcal{U}, \quad \mathbf{w} \in \mathcal{U}^\perp. \quad (1.2.11)$$

Der eindeutig bestimmte Vektor  $\mathbf{u}_0 \in \mathcal{U}$  mit  $\mathbf{v} - \mathbf{u}_0 \in \mathcal{U}^\perp$  ist auch der eindeutig bestimmte Vektor aus  $\mathcal{U}$  mit dem kleinsten Abstand von  $\mathbf{v}$ :

$$\|\mathbf{v} - \mathbf{u}_0\| = \min_{\mathbf{u} \in \mathcal{U}} \|\mathbf{v} - \mathbf{u}\|. \quad (1.2.12)$$

**Beweis:**

1. Es sei  $\mathbf{v} \in \mathcal{V}$  und  $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$  eine Orthogonalbasis von  $\mathcal{U}$ . Ein Vektor

$$\mathbf{u}_0 = \sum_{i=1}^m c_i \mathbf{a}_i \quad (1.2.13)$$

aus  $\mathcal{U}$  erfüllt

$$(\mathbf{v} - \mathbf{u}_0) \in \mathcal{U}^\perp \quad (1.2.14)$$

aufgrund von Satz 1.2.2 (5.) genau dann, wenn  $\mathbf{v} - \mathbf{u}_0$  zu jedem Basisvektor  $\mathbf{a}_i$  orthogonal ist, das heißt, genau dann wenn

$$\langle \mathbf{a}_i, \mathbf{v} \rangle = \langle \mathbf{a}_i, \mathbf{u}_0 \rangle = \left\langle \mathbf{a}_i, \sum_{j=1}^m c_j \mathbf{a}_j \right\rangle = \sum_{j=1}^m c_j \langle \mathbf{a}_i, \mathbf{a}_j \rangle = c_i \langle \mathbf{a}_i, \mathbf{a}_i \rangle \quad (1.2.15)$$

für  $1 \leq i \leq m$ . Somit erfüllt der Vektor

$$\mathbf{u}_0 = \sum_{i=1}^m \frac{\langle \mathbf{a}_i, \mathbf{v} \rangle}{\langle \mathbf{a}_i, \mathbf{a}_i \rangle} \mathbf{a}_i \quad (1.2.16)$$

und nur dieser aus  $\mathcal{U}$  die Bedingung (1.2.14).

2. Es sei  $\mathbf{v} \in \mathcal{V}$ ,  $\mathbf{u} \in \mathcal{U}$  und  $\mathbf{u}_0$  der nach Teil 1 des Satzes existierende und eindeutig bestimmte Vektor aus  $\mathcal{U}$  mit  $(\mathbf{v} - \mathbf{u}_0) \in \mathcal{U}^\perp$ . Dann ergibt sich mit dem Satz von Pythagoras, da  $(\mathbf{v} - \mathbf{u}_0) \perp (\mathbf{u}_0 - \mathbf{u})$ :

$$\begin{aligned} \|\mathbf{v} - \mathbf{u}\|^2 &= \|(\mathbf{v} - \mathbf{u}_0) + (\mathbf{u}_0 - \mathbf{u})\|^2 \\ &= \|\mathbf{v} - \mathbf{u}_0\|^2 + \|\mathbf{u}_0 - \mathbf{u}\|^2 \\ &\geq \|\mathbf{v} - \mathbf{u}_0\|^2 \\ „ = “ &\Leftrightarrow \mathbf{u} = \mathbf{u}_0 \end{aligned} \quad (1.2.17)$$

**Definition 1.2.3.**

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt,  $\mathcal{U}$  ein Untervektorraum von  $\mathcal{V}$ . Für  $\mathbf{v} \in \mathcal{V}$  sei  $\mathbf{u}(\mathbf{v})$  der eindeutig bestimmte Vektor aus  $\mathcal{U}$  mit  $(\mathbf{v} - \mathbf{u}(\mathbf{v})) \in \mathcal{U}^\perp$ . Dann heißt die Abbildung

$$p_{\mathcal{U}} : \begin{cases} \mathcal{V} \rightarrow \mathcal{U} \\ \mathbf{v} \mapsto p_{\mathcal{U}}(\mathbf{v}) = \mathbf{u}(\mathbf{v}) \end{cases} \quad (1.2.18)$$

Orthogonalprojektion von  $\mathcal{V}$  auf  $\mathcal{U}$ .

**Korollar 1.**

1. Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt,  $\mathcal{U}$  ein Untervektorraum von  $\mathcal{V}$  und  $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$  eine Orthogonalbasis von  $\mathcal{U}$ . Dann ist

$$p_{\mathcal{U}}(\mathbf{v}) = \sum_{i=1}^m \frac{\langle \mathbf{a}_i, \mathbf{v} \rangle}{\langle \mathbf{a}_i, \mathbf{a}_i \rangle} \mathbf{a}_i = \sum_{i=1}^m p_{\mathcal{L}(\{\mathbf{a}_i\})}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathcal{V} \quad (1.2.19)$$

die Orthogonalprojektion von  $\mathcal{V}$  auf  $\mathcal{U}$ .

2. Es sei  $\mathcal{U}$  ein Untervektorraum des  $\mathbb{R}^n$  mit dem kanonischen Skalarprodukt,  $\{\vec{q}_1, \dots, \vec{q}_m\}$  eine Orthonormalbasis von  $\mathcal{U}$  und  $Q = [\vec{q}_1 \quad \vec{q}_2 \quad \dots \quad \vec{q}_m]$  die aus den Basisvektoren gebildete Matrix. Dann ist

$$p_{\mathcal{U}}(\vec{y}) = \sum_{i=1}^m (\vec{q}_i^T \vec{y}) \vec{q}_i = \sum_{i=1}^m z_i \vec{q}_i = Q \vec{z} = Q Q^T \vec{y} \quad \forall \vec{y} \in \mathbb{R}^n \quad (1.2.20)$$

die Orthogonalprojektion des  $\mathbb{R}^n$  auf  $\mathcal{U}$ . Hierbei ist  $z_i = \vec{q}_i^T \vec{y}$  und  $\vec{z} = (z_1, \dots, z_m)^T = Q^T \vec{y}$  der Koordinatenvektor der Orthogonalprojektion von  $\vec{y}$  bezüglich der Basis  $\{\vec{q}_1, \dots, \vec{q}_m\}$ .

**Korollar 2. Cauchy-Schwarzsche Ungleichung**

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt und  $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ . Dann ergibt sich mit dem Satz von Pythagoras, da  $\mathbf{v} - p_{\mathcal{L}(\{\mathbf{u}\})}(\mathbf{v}) \perp p_{\mathcal{L}(\{\mathbf{u}\})}(\mathbf{v})$ :

$$\begin{aligned} \|\mathbf{v}\|^2 &= \|\mathbf{v} - p_{\mathcal{L}(\{\mathbf{u}\})}(\mathbf{v})\|^2 + \|p_{\mathcal{L}(\{\mathbf{u}\})}(\mathbf{v})\|^2 \\ &\geq \|p_{\mathcal{L}(\{\mathbf{u}\})}(\mathbf{v})\|^2 = \left\| \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u} \right\|^2 = \frac{\langle \mathbf{u}, \mathbf{v} \rangle^2}{\langle \mathbf{u}, \mathbf{u} \rangle} \end{aligned} \quad (1.2.21)$$

Hieraus ergibt sich unmittelbar die Cauchy-Schwarzsche Ungleichung

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \quad (1.2.22)$$

Das Gleichheitszeichen gilt in Gleichung (1.2.21) und somit auch in Gleichung (1.2.22) genau dann, wenn  $\mathbf{v} = p_{\mathcal{L}(\{\mathbf{u}\})}(\mathbf{v})$ , also genau dann, wenn  $\mathbf{v} = c\mathbf{u}$  mit  $c \in \mathbb{R}$ .



**Satz und Definition 1.2.4.**

Es seien  $\mathcal{U}, \mathcal{W}$  Untervektorräume des Vektorraums  $\mathcal{V}$  und  $\mathbf{o}$  der Nullvektor.

1. Der Vektorraum

$$\mathcal{U} + \mathcal{W} := \mathcal{L}(\mathcal{U} \cup \mathcal{W}) = \{\mathbf{v} \mid \mathbf{v} = \mathbf{u} + \mathbf{w}, \mathbf{u} \in \mathcal{U}, \mathbf{w} \in \mathcal{W}\} \quad (1.2.23)$$

heißt Summe der Vektorräume  $\mathcal{U}$  und  $\mathcal{W}$ .

2. Die Vektorräume  $\mathcal{U}$  und  $\mathcal{W}$  heißen linear unabhängig genau dann, wenn

$$\mathcal{U} \cap \mathcal{W} = \{\mathbf{o}\} \quad (1.2.24)$$

3. Die Vektorräume  $\mathcal{U}$  und  $\mathcal{W}$  sind linear unabhängig genau dann, wenn jeder Vektor  $\mathbf{v}$  aus  $\mathcal{V}$  höchstens eine Darstellung der Form  $\mathbf{v} = \mathbf{u} + \mathbf{w}$  mit  $\mathbf{u} \in \mathcal{U}$  und  $\mathbf{w} \in \mathcal{W}$  besitzt.

4. Die Summe linear unabhängiger Untervektorräume  $\mathcal{U}$  und  $\mathcal{W}$  heißt direkt und wird mit  $\mathcal{U} \oplus \mathcal{W}$  bezeichnet.

5. Ist  $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$  und  $\mathbf{v} = \mathbf{u} + \mathbf{w} \in \mathcal{V}$  mit  $\mathbf{u} \in \mathcal{U}, \mathbf{w} \in \mathcal{W}$ , so nennt man  $\mathcal{W}$  ein Komplement von  $\mathcal{U}$  in  $\mathcal{V}$  und  $\mathbf{u}$  die Komponente von  $\mathbf{v}$  in  $\mathcal{U}$  in bezug auf die direkte Zerlegung von  $\mathcal{V}$  in  $\mathcal{U}$  und  $\mathcal{W}$ . Die lineare Abbildung  $p_{\mathcal{U}}^{\mathcal{V}}$  auf  $\mathcal{V}$ , die jedem Vektor  $\mathbf{v}$  aus  $\mathcal{V}$  seine Komponente in  $\mathcal{U}$  zuordnet, heißt Projektion längs  $\mathcal{W}$  von  $\mathcal{V}$  auf  $\mathcal{U}$ .

6. Ist  $\mathcal{W}$  ein Komplement von  $\mathcal{U}$  in  $\mathcal{V}$ , dann gelten die folgenden Aussagen:

$$p_{\mathcal{W}}^{\mathcal{U}} = id_{\mathcal{V}} - p_{\mathcal{U}}^{\mathcal{W}} \quad (1.2.25)$$

$$p_{\mathcal{U}}^{\mathcal{W}}(\mathbf{u}) = \mathbf{u} \quad \forall \mathbf{u} \in \mathcal{U} \quad (1.2.26)$$

$$\mathcal{U} \text{ ist das Bild und } \mathcal{W} \text{ der Kern von } p_{\mathcal{U}}^{\mathcal{W}} \quad (1.2.27)$$

Für  $\mathcal{V}$  endlich-dimensional gilt außerdem

$$\dim(\mathcal{V}) = \dim(\mathcal{U}) + \dim(\mathcal{W}) \quad (1.2.28)$$

**Satz 1.2.5.**

Es sei  $\mathcal{V}$  ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt,  $\mathcal{U}$  ein Untervektorraum von  $\mathcal{V}$ . Dann ist  $\mathcal{U}^{\perp}$  ein Vektorraum mit

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp} \quad (1.2.29)$$

$$(\mathcal{U}^{\perp})^{\perp} = \mathcal{U} \quad (1.2.30)$$

$$p_{\mathcal{U}} = p_{\mathcal{U}^{\perp}}^{\perp} \quad (1.2.31)$$

$\mathcal{U}^{\perp}$  heißt orthogonales Komplement von  $\mathcal{U}$ .

**Definition 1.2.4.**

Eine quadratische reelle Matrix heißt idempotent genau dann, wenn

$$A^2 = A. \quad (1.2.32)$$

**Satz 1.2.6.**

1. Eine quadratische reelle Matrix  $P$  der Ordnung  $n$  ist die Matrix einer Projektion des  $\mathbb{R}^n$  genau dann, wenn  $P$  idempotent ist.
2. Eine quadratische reelle Matrix  $P$  der Ordnung  $n$  ist die Matrix einer Orthogonalprojektion des  $\mathbb{R}^n$  mit dem kanonischen Skalarprodukt genau dann, wenn  $P$  symmetrisch und idempotent ist.
3. Ist  $P$  eine symmetrische und idempotente reelle Matrix der Ordnung  $n$ , dann ist  $P$  die Matrix der Orthogonalprojektion des  $\mathbb{R}^n$  mit dem kanonischen Skalarprodukt auf das Bild bzw. den Spaltenraum  $\mathfrak{Im}(P)$  von  $P$ .  $I - P$  ist die Matrix der Orthogonalprojektion auf den Kern  $\mathfrak{Im}(P)^\perp$  von  $P$ .

**Beweis:**

2. Es sei  $P$  die Matrix einer Orthogonalprojektion. Aus Korollar 1.2 folgt, dass es eine Matrix  $Q$  mit  $Q^T Q = I$  und  $P = Q Q^T$  gibt. Aus dieser Darstellung folgt unmittelbar die Symmetrie und Idempotenz von  $P$ .

Es sei  $P$  eine symmetrische und idempotente Matrix.

Für  $\vec{y} \in \mathbb{R}^n$  gilt

$$\vec{y} = P\vec{y} + (\vec{y} - P\vec{y}), \quad P\vec{y} \in \mathfrak{Im}(P) \quad (1.2.33)$$

Für alle  $\vec{x} \in \mathbb{R}^n$  gilt

$$\begin{aligned} (\vec{y} - P\vec{y})^T P\vec{x} &= \vec{y}^T P\vec{x} - \vec{y}^T P^T P\vec{x} \\ &= \vec{y}^T P\vec{x} - \vec{y}^T P P\vec{x} \quad (\text{da } P \text{ symmetrisch}) \\ &= \vec{y}^T P\vec{x} - \vec{y}^T P\vec{x} \quad (\text{da } P \text{ idempotent}) \\ &= 0, \end{aligned} \quad (1.2.34)$$

das heißt  $(\vec{y} - P\vec{y}) \in \mathfrak{Im}(P)^\perp$ . Somit ist  $P\vec{y}$  die Orthogonalprojektion von  $\vec{y}$  auf  $\mathfrak{Im}(P)$ .

**Satz 1.2.7.**

Es sei  $\mathcal{V}$  ein  $n$ -dimensionaler reeller Vektorraum mit Skalarprodukt,  $\mathcal{U}$  ein  $r$ -dimensionaler Untervektorraum von  $\mathcal{V}$ ,  $\mathcal{S} = \{\mathfrak{x}_1, \dots, \mathfrak{x}_k\}$  ein Erzeugendensystem von  $\mathcal{U}$ . Die orthogonale Projektion eines Vektors  $\mathfrak{v} \in \mathcal{V}$  auf  $\mathcal{U}$  ist der eindeutig bestimmte Vektor

$$p_{\mathcal{U}}(\mathfrak{v}) = \sum_{j=1}^k b_j \mathfrak{x}_j \quad (1.2.35)$$

aus  $\mathcal{U}$  mit

$$(\mathfrak{v} - p_{\mathcal{U}}(\mathfrak{v})) \in \mathcal{U}^\perp = \mathcal{S}^\perp \quad (1.2.36)$$

Diese Bedingung ist äquivalent zu den Bedingungen

$$(\mathbf{v} - p_{\mathcal{U}}(\mathbf{v})) \perp \mathbf{x}_i \quad \text{für } i = 1, \dots, k \quad (1.2.37)$$

die wiederum genau dann erfüllt sind, wenn die Koeffizienten  $b_1, \dots, b_k$  die (nicht notwendig eindeutig lösaren) sogenannten Normalgleichungen erfüllen

$$\sum_{j=1}^k \langle \mathbf{x}_i, \mathbf{x}_j \rangle b_j = \langle \mathbf{x}_i, \mathbf{v} \rangle \quad \text{für } i = 1, \dots, k \quad (1.2.38)$$

In Matrizenform lauten die Normalgleichungen

$$M\vec{b} = \vec{m} \quad (1.2.39)$$

mit  $M = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle)_{i,j=1,\dots,k}$ ,  $\vec{b} = (b_1, \dots, b_k)^T$  und  $\vec{m} = (\langle \mathbf{x}_1, \mathbf{v} \rangle, \dots, \langle \mathbf{x}_k, \mathbf{v} \rangle)^T$ .

### Korollar 1.

Es sei

$$X = (x_{ij})_{i=1,\dots,n, j=1,\dots,k} = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \dots & \vec{x}_k \end{bmatrix} \quad (1.2.40)$$

eine reelle  $(n \times k)$ -Matrix mit dem Spaltenraum  $\mathfrak{Im}(X) = \mathcal{L}(\{\vec{x}_1, \dots, \vec{x}_k\})$ . Die Matrix der Orthogonalprojektion des  $\mathbb{R}^n$  mit dem kanonischen Skalarprodukt auf den Spaltenraum von  $X$  wird im Folgenden mit  $P_{\mathfrak{Im}(X)}$  bezeichnet.

Für zwei beliebige Vektoren  $\vec{y} \in \mathbb{R}^n$  und  $\vec{b} = (b_1, \dots, b_k)^T \in \mathbb{R}^k$  sind die folgenden Aussagen äquivalent:

1.  $X\vec{b}$  ist die orthogonale Projektion von  $\vec{y}$  auf den Spaltenraum der Matrix  $X$ :

$$P_{\mathfrak{Im}(X)}\vec{y} = \sum_{j=1}^k \vec{x}_j b_j = X\vec{b}. \quad (1.2.41)$$

2.  $\vec{b}$  ist Lösung der Minimierungsaufgabe

$$\|\vec{y} - X\vec{b}\| = \min_{\mathbf{u} \in \mathfrak{Im}(X)} \|\vec{y} - \mathbf{u}\| = \min_{\vec{\beta} \in \mathbb{R}^k} \|\vec{y} - X\vec{\beta}\|. \quad (1.2.42)$$

3.  $\vec{b}$  ist Lösung der Normalgleichungen:

$$X^T X \vec{b} = X^T \vec{y} \quad (1.2.43)$$

**Satz 1.2.8.**

Es sei  $A$  eine  $(n \times m)$ -Matrix mit den Spaltenvektoren  $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$

1. Es sei  $\mathfrak{I}\mathfrak{m}(A) = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^m\} = \mathcal{L}(\{\vec{a}_1, \dots, \vec{a}_m\})$  das Bild bzw. der Spaltenraum von  $A$  und  $\mathfrak{K}(A) = \{\vec{x} \in \mathbb{R}^m \mid A\vec{x} = \vec{0}\}$  der Kern von  $A$ , dann gilt:

$$rg(A) = \dim(\mathfrak{I}\mathfrak{m}(A)) = \dim(\mathfrak{I}\mathfrak{m}(A^T)) = m - \dim(\mathfrak{K}(A)) \quad (1.2.44)$$

2. Es sei  $B$  eine  $(m \times k)$ -Matrix. Dann gilt:

$$rg(AB) \leq \min\{rg(A), rg(B)\} \quad (1.2.45)$$

3. Es gilt:

$$rg(A^T A) = rg(A) = rg(A^T) = rg(AA^T) \quad (1.2.46)$$

$$\mathfrak{I}\mathfrak{m}(A^T) = \mathfrak{I}\mathfrak{m}(A^T A) \quad (1.2.47)$$

$A^T A$  und  $AA^T$  sind symmetrisch und positiv-semidefinit.

$A^T A$  (bzw.  $AA^T$ ) ist regulär und somit positiv-definit genau dann, wenn  $rg(A) = m$  (bzw.  $rg(A) = n$ ).

**Bemerkung 1. Motivation der Definition der verallgemeinerten Inversen**

Es seien  $\mathcal{V}$  und  $\mathcal{W}$  endlich-dimensionale Vektorräume und  $l$  eine lineare Abbildung von  $\mathcal{V}$  in  $\mathcal{W}$ .

1. Es sei  $\mathcal{U}$  ein Komplement des Kerns  $\mathfrak{K}(l)$ . Dann ist  $l$  eingeschränkt auf  $\mathcal{U}$  injektiv und  $l(\mathcal{U}) = \mathfrak{I}\mathfrak{m}(l)$ , das heißt  $\mathcal{U}$  und  $\mathfrak{I}\mathfrak{m}(l)$  sind isomorph.
2. Es sei  $\mathcal{U}$  ein Komplement des Kerns  $\mathfrak{K}(l)$  und  $\tilde{l}$  die bijektive Abbildung

$$\tilde{l} : \begin{cases} \mathcal{U} \rightarrow \mathfrak{I}\mathfrak{m}(l) \\ \mathfrak{x} \mapsto l(\mathfrak{x}) \end{cases} \quad (1.2.48)$$

Dann existiert eine Umkehrabbildung  $\tilde{l}^{-1}$  von  $\tilde{l}$ . Für jede lineare Erweiterung  $l^{-}$  von  $\tilde{l}^{-1}$  auf den Definitionsbereich  $\mathcal{W}$  und den Wertevorrat  $\mathcal{V}$  gilt:

$$l \circ l^{-} \circ l = l \quad (1.2.49)$$

Die lineare Abbildung  $l^{-}$  bildet jeden Bildpunkt auf ein Urbild ab.

Ist  $A$  eine Abbildungsmatrix von  $l$  und  $B$  von  $l^{-}$ , so gilt  $ABA = A$

**Definition 1.2.5.**

Es sei  $A$  eine  $(n \times m)$ -Matrix. Eine  $(m \times n)$ -Matrix  $A^{-}$  heißt verallgemeinerte Inverse oder Pseudoinverse von  $A$ , wenn

$$A = AA^{-}A \quad (1.2.50)$$

gilt.

**Bemerkung 1.**

Ist das Gleichungssystem  $A\vec{x} = \vec{y}$  konsistent und  $A^-$  eine verallgemeinerte Inverse von  $A$ , dann ist  $\vec{x}_0 = A^- \vec{y}$  eine Lösung des Gleichungssystems.

**Satz 1.2.9.**

Es sei  $A$  eine Matrix.

1. Es existiert eine verallgemeinerte Inverse  $A^-$  von  $A$ , die im Allgemeinen nicht eindeutig ist.
2. Es sei  $A^-$  eine verallgemeinerte Inverse von  $A$ , dann ist  $(A^-)^T$  eine verallgemeinerte Inverse von  $A^T$ .

**Satz 1.2.10.**

Es sei  $X$  eine reelle  $(n \times k)$ -Matrix. Dann gelten für jede verallgemeinerte Inverse  $(X^T X)^-$  von  $(X^T X)$  die folgenden Aussagen:

1. Der Vektor

$$\vec{b}_0 = (X^T X)^- X^T \vec{y} \quad (1.2.51)$$

ist eine Lösung der Normalgleichungen aus Korollar 1 zu Satz 1.2.7:

$$X^T X \vec{b} = X^T \vec{y} \quad (1.2.52)$$

2. Die Matrix der Orthogonalprojektion auf den Spaltenraum von  $X$  ist

$$P_{\mathcal{M}(X)} = X(X^T X)^- X^T \quad (1.2.53)$$

3. Die Matrix  $X(X^T X)^-$  ist eine verallgemeinerte Inverse von  $X^T$ .

4. Die Matrix  $(X^T X)^- X^T$  ist eine verallgemeinerte Inverse von  $X$ .

**Beweis:**

1. Die Aussage folgt unmittelbar aus Bemerkung 1 zu Definition 1.2.5.

2. Mit Korollar 1 zu Satz 1.2.7 und Teil 1 dieses Satzes ergibt sich

$$P_{\mathcal{M}(X)} \vec{y} = X \vec{b}_0 = X(X^T X)^- X^T \vec{y} \quad \forall \vec{y} \in \mathbb{R}^n \quad (1.2.54)$$

hieraus folgt

$$P_{\mathcal{M}(X)} = X(X^T X)^- X^T \quad (1.2.55)$$

3. Aus 1. folgt

$$X^T \vec{y} = X^T X \vec{b}_0 = X^T X(X^T X)^- X^T \vec{y} \quad \forall \vec{y} \in \mathbb{R}^n \quad (1.2.56)$$

hieraus folgt

$$X^T X(X^T X)^- X^T = X^T \quad (1.2.57)$$

### 1.3 Wahrscheinlichkeitsrechnung

#### Mehrdimensionale stetige Veteilungen

**Definition 1.3.1. Mehrdimensionale Zufallsvariable, Zufallsvektor**

Es seien  $X_1, \dots, X_n$  Zufallsvariablen über dem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$ .

1. Die Abbildung

$$\vec{X} := (X_1, \dots, X_n)^T : \Omega \rightarrow \mathbb{R}^n, \quad \vec{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))^T \quad (1.3.1)$$

heißt  $n$ -dimensionaler Zufallsvektor,  $n$ -dimensionale Zufallsvariable oder  $n$ -dimensionale Zufallsgröße über  $(\Omega, \mathfrak{A}, P)$ .

2. Die Funktion  $F_{\vec{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$  mit

$$\begin{aligned} F_{\vec{X}}(\vec{x}) &= F_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n) = P(\vec{X} \leq \vec{x}) \\ &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \quad \forall \vec{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n \end{aligned} \quad (1.3.2)$$

heißt Verteilungsfunktion des Zufallsvektors  $\vec{X}$  oder gemeinsame Verteilungsfunktion der Zufallsvariablen  $X_1, \dots, X_n$ .

3. Die Zufallsvariablen  $X_1, \dots, X_n$  mit den Verteilungsfunktionen  $F_{X_1}, \dots, F_{X_n}$  heißen (stochastisch) unabhängig, falls für die gemeinsame Verteilungsfunktion  $F_{X_1, \dots, X_n}$  von  $X_1, \dots, X_n$  gilt:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R} \quad (1.3.3)$$

$X_1, \dots, X_n$  heißen paarweise unabhängig  $:\Leftrightarrow X_i, X_j$  unabhängig  $\forall i \neq j$ .

4. Sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein Zufallsvektor mit der Verteilungsfunktion  $F_{\vec{X}}$ . Existiert eine nicht-negative Funktion  $f_{\vec{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$  mit

$$F_{\vec{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\vec{X}}(\xi_1, \dots, \xi_n) d\xi_n \dots d\xi_1, \quad (1.3.4)$$

dann heißt  $\vec{X}$  stetige (stetig verteilte)  $n$ -dimensionale Zufallsvariable oder stetiger (stetig verteilter) Zufallsvektor.

$f_{\vec{X}}$  ist die (Verteilungs-)Dichte von  $\vec{X}$  bzw. die gemeinsame Dichte von  $X_1, \dots, X_n$ .

**Satz 1.3.1. Marginale Verteilung**

Es sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein Zufallsvektor mit der Verteilungsfunktion  $F_{\vec{X}}$ .

Die Verteilung eines Teilvektors  $(X_{i_1}, \dots, X_{i_r})^T$  ( $1 \leq i_1 < i_2 < \dots < i_r \leq n$ ;  $1 \leq r < n$ ) bezeichnet man als marginale Verteilung oder Randverteilung von  $(X_{i_1}, \dots, X_{i_r})^T$ .

1. Die (marginale) Verteilungsfunktion  $F_{X_{i_1}, \dots, X_{i_r}}$  von  $(X_{i_1}, \dots, X_{i_r})^T$  ergibt sich dann aus der Beziehung

$$F_{X_{i_1}, \dots, X_{i_r}}(x_{i_1}, \dots, x_{i_r}) = F_{\vec{X}}(\underbrace{\infty, \dots, \infty}_{i_1-1}, \underbrace{\infty, \dots, \infty}_{i_2-i_1-1}, x_{i_2}, \dots, x_{i_r}, \underbrace{\infty, \dots, \infty}_{n-i_r}). \quad (1.3.5)$$

2. Ist  $\vec{X} = (X_1, \dots, X_n)^T$  stetig verteilt mit der Dichte  $f_{\vec{X}}$ , dann ist auch jeder Teilvektor stetig verteilt. Die marginale Dichte bzw. Randdichte von  $(X_1, \dots, X_r)^T$  ergibt sich zu

$$f_{X_1, \dots, X_r}(x_1, \dots, x_r) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\vec{X}}(x_1, \dots, x_r, \xi_{r+1}, \dots, \xi_n) d\xi_{r+1} \dots d\xi_n \quad (1.3.6)$$

Eine entsprechende Aussage gilt für die Randdichte von  $(X_{i_1}, \dots, X_{i_r})$  ( $1 \leq i_1 < i_2 < \dots < i_r \leq n, 1 \leq r < n$ ).

### Definition 1.3.2. Unabhängigkeit von Zufallsvektoren

Der  $n$ -dimensionale Zufallsvektor  $\vec{X} = [\vec{X}_1^T \dots \vec{X}_k^T]^T$  sei in  $k$  Teilvektoren  $\vec{X}_1, \dots, \vec{X}_k$  partitioniert. Die Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_k$  mit den Verteilungsfunktionen  $F_{\vec{X}_1}, \dots, F_{\vec{X}_k}$  heißen (stochastisch) unabhängig, falls für die gemeinsame Verteilungsfunktion  $F_{\vec{X}}$  gilt:

$$F_{\vec{X}}(\vec{x}) = F_{\vec{X}_1, \dots, \vec{X}_k}(\vec{x}_1, \dots, \vec{x}_k) = F_{\vec{X}_1}(\vec{x}_1) \cdot \dots \cdot F_{\vec{X}_k}(\vec{x}_k) \quad \forall \vec{x} = [\vec{x}_1^T \dots \vec{x}_k^T]^T \in \mathbb{R}^n. \quad (1.3.7)$$

### Satz 1.3.2. Unabhängige Zufallsvektoren

Der  $n$ -dimensionale Zufallsvektor  $\vec{X} = [\vec{X}_1^T \dots \vec{X}_k^T]^T$  sei in  $k$  Teilvektoren  $\vec{X}_1, \dots, \vec{X}_k$  partitioniert.

1. Die Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_k$  seien stochastisch unabhängig und  $g_1, \dots, g_k$  hinreichend reguläre Abbildungen (z.B. stetig, genau: Lebesgue-messbar). Dann sind auch  $\vec{Y}_1 = g_1(\vec{X}_1), \dots, \vec{Y}_k = g_k(\vec{X}_k)$  stochastisch unabhängig.
2. Die Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_k$  seien stetig verteilt. Dann sind sie stochastisch unabhängig genau dann, wenn das Produkt der marginalen Dichten

$$f(\vec{x}_1, \dots, \vec{x}_n) = f_{\vec{X}_1}(\vec{x}_1) \cdot f_{\vec{X}_2}(\vec{x}_2) \cdot \dots \cdot f_{\vec{X}_k}(\vec{x}_k) \quad \forall \vec{x} = [\vec{x}_1^T \dots \vec{x}_k^T]^T \in \mathbb{R}^n \quad (1.3.8)$$

eine Dichte von  $\vec{X}$  und somit von der gemeinsamen Verteilung von  $\vec{X}_1, \dots, \vec{X}_k$  ist.

### Satz 1.3.3. Transformationssatz für Dichten

Es sei  $\vec{X}$  ein stetiger  $n$ -dimensionaler Zufallsvektor mit der Dichte  $f_{\vec{X}}(x_1, \dots, x_n)$ ,  $A$  eine offene Teilmenge des  $\mathbb{R}^n$  mit  $P(\vec{X} \in A) = 1$  und  $g : A \rightarrow \mathbb{R}^n$  eine auf  $A$  injektive stetig differenzierbare Abbildung mit einer auf ganz  $A$  regulären Funktionalmatrix. Dann ist auch der Zufallsvektor  $\vec{Y} = g(\vec{X})$  stetig verteilt mit der Dichte

$$f_{\vec{Y}}(\vec{y}) = \begin{cases} f_{\vec{X}}(g^{-1}(\vec{y})) \left| \det \left( \frac{\partial g^{-1}}{\partial \vec{y}}(\vec{y}) \right) \right| & \text{für } \vec{y} \in g(A) \\ 0 & \text{für } \vec{y} \notin g(A) \end{cases} \quad (1.3.9)$$

### Beweis:

Der Beweis folgt aus dem Satz über die Umkehrabbildung stetig differenzierbarer Abbildungen und dem Satz über die Variablentransformation bei mehrfachen Integralen aus der Analysis.

**Definition 1.3.3. Erwartungswert**

Es sei  $\vec{X}$  ein stetiger  $n$ -dimensionaler Zufallsvektor mit der Dichte  $f_{\vec{X}}$  und  $Y = g(\vec{X})$  eine transformierte Zufallsvariable. Dann heißt

$$\begin{aligned}\mu_Y = E[Y] &= E[g(\vec{X})] = \int_{\mathbb{R}^n} g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x} \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) \cdot f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n,\end{aligned}\quad (1.3.10)$$

falls  $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1, \dots, x_n)| \cdot f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n$  endlich ist, Erwartungswert von  $Y = g(\vec{X})$ .

**Satz 1.3.4. Erwartungswert des Produktes unabhängiger Zufallsgrößen**

Es seien  $X_1, \dots, X_n$  unabhängige Zufallsgrößen mit existierenden Erwartungswerten, dann gilt

$$E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n E[X_i] \quad (1.3.11)$$

**Satz und Definition 1.3.5. Momentenerzeugende Funktion**

1. Es sei  $\vec{X}$  ein  $n$ -dimensionaler Zufallsvektor. Existiert die Funktion

$$m_{\vec{X}}(\vec{t}) = E[e^{\vec{t}^T \vec{X}}] \quad (1.3.12)$$

für  $-h < t_i < h$ ,  $i = 1, \dots, n$ , für ein  $h > 0$ , dann heißt diese Funktion momentenerzeugende Funktion von  $\vec{X}$ .

2. Es seien  $\vec{X}$  und  $\vec{Y}$   $n$ -dimensionale Zufallsvektoren mit existierenden momentenerzeugenden Funktionen. Sie haben die gleiche Verteilung genau dann, wenn ihre momentenerzeugenden Funktionen auf einem offenen den Ursprung enthaltenen  $n$ -dimensionalen Rechteck übereinstimmen.

3. Der  $n$ -dimensionale Zufallsvektor  $\vec{X} = [\vec{X}_1^T \dots \vec{X}_k^T]^T$  sei in  $k$  Teilvektoren  $\vec{X}_1, \dots, \vec{X}_k$  partitioniert, deren momentenerzeugende Funktionen  $m_{\vec{X}_1}, \dots, m_{\vec{X}_k}$  existieren. Die Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_k$  sind stochastisch unabhängig genau dann, wenn

$$m_{\vec{X}}(\vec{t}) = \prod_{i=1}^k m_{\vec{X}_i}(\vec{t}_i), \quad \vec{t} = [\vec{t}_1^T \dots \vec{t}_k^T]^T \quad (1.3.13)$$

auf einem offenen den Ursprung enthaltenen  $n$ -dimensionalen Rechteck gilt.



**Beweis:**

Beweis von 3.

Sind die Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_k$  stochastisch unabhängig, dann gilt

$$\begin{aligned} m_{\vec{X}}(\vec{t}) &= E[e^{\vec{t}^T \vec{X}}] = E[e^{\sum_{i=1}^k \vec{t}_i^T \vec{X}_i}] = E\left[\prod_{i=1}^k e^{\vec{t}_i^T \vec{X}_i}\right] \\ &= \prod_{i=1}^k E[e^{\vec{t}_i^T \vec{X}_i}] = \prod_{i=1}^k m_{\vec{X}_i}(\vec{t}_i) \end{aligned} \quad (1.3.14)$$

Ist die Gleichung (1.3.13) erfüllt, dann stimmt aufgrund von Gleichung (1.3.14) die momentenerzeugende Funktion der gemeinsamen Verteilung der Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_k$  mit der momentenerzeugenden Funktion der gemeinsamen Verteilungsfunktion überein, die bei Unabhängigkeit der Zufallsvektoren vorliegt. Hieraus folgt mit 2. die Unabhängigkeit der Zufallsvektoren.

**Definition 1.3.4. Kovarianzmatrix und Zufallsmatrix**

1. Seien  $X, Y$  Zufallsvariablen über einem gemeinsamen Wahrscheinlichkeitsraum, dann heißt

$$\sigma_{X,Y} = \text{Kov}(X, Y) := E[(X - E[X])(Y - E[Y])] \quad (1.3.15)$$

Kovarianz von  $X$  und  $Y$ , und

$$\sigma_X^2 = \text{Var}(X) := \sigma_{X,X} = \text{Kov}(X, X) = E[(X - E[X])^2] \quad (1.3.16)$$

heißt Varianz von  $X$ . Existieren die zweiten Momente von  $X$  und  $Y$ , dann heißt

$$\rho_{X,Y} := \begin{cases} \frac{\text{Kov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} & \text{falls } \text{Var}(X) > 0, \text{Var}(Y) > 0 \\ 0 & \text{falls } \text{Var}(X) = 0 \text{ oder } \text{Var}(Y) = 0 \end{cases} \quad (1.3.17)$$

Korrelationskoeffizient von  $X$  und  $Y$ , und  $\rho_{X,Y}^2$  heißt Bestimmtheitsmaß.

2. Es sei

$$\mathcal{X} = (X_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \quad (1.3.18)$$

eine Zufallsmatrix, das heißt, eine Matrix von Zufallsvariablen über einem gemeinsamen Wahrscheinlichkeitsraum. Existieren die Erwartungswerte der Komponenten von  $\mathcal{X}$ , dann heißt die Matrix

$$E[\mathcal{X}] = (E[X_{ij}])_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \quad (1.3.19)$$

Erwartungswert oder Erwartungswertmatrix von  $\mathcal{X}$ .

3. Es sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein  $n$ -dimensionaler Zufallsvektor. Existieren die Erwartungswerte der Komponenten von  $\vec{X}$ , dann heißt

$$\vec{\mu}_{\vec{X}} = E[\vec{X}] = (E[X_1], \dots, E[X_n])^T \quad (1.3.20)$$

Erwartungswert oder Erwartungswertvektor von  $\vec{X}$ .

4. Es sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein  $n$ -dimensionaler Zufallsvektor und  $\vec{Y} = (Y_1, \dots, Y_m)^T$  ein  $m$ -dimensionaler Zufallsvektor. Existieren die zweiten Momente der Komponenten von  $\vec{X}$  und  $\vec{Y}$ , dann heißt

$$\begin{aligned} \Sigma_{\vec{X}, \vec{Y}} &= \text{Kov}(\vec{X}, \vec{Y}) = \left( \text{Kov}(X_i, Y_j) \right)_{\substack{i=1, \dots, n \\ j=1, \dots, m}} = E \left[ \left( (X_i - E[X_i])(Y_j - E[Y_j]) \right)_{\substack{i=1, \dots, n \\ j=1, \dots, m}} \right] \\ &= E \left[ (\vec{X} - E[\vec{X}])(\vec{Y} - E[\vec{Y}])^T \right] \end{aligned} \quad (1.3.21)$$

Kovarianzmatrix von  $\vec{X}$  und  $\vec{Y}$ . Die Vektoren heißen unkorreliert genau dann, wenn  $\Sigma_{\vec{X}, \vec{Y}} = O$

5. Es sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein  $n$ -dimensionaler Zufallsvektor. Existieren die zweiten Momente der Komponenten von  $\vec{X}$ , dann heißt

$$\begin{aligned} \Sigma_{\vec{X}} &= \text{Kov}(\vec{X}) = \text{Kov}(\vec{X}, \vec{X}) = E \left[ \left( (X_i - E[X_i])(X_j - E[X_j]) \right)_{\substack{i=1, \dots, n \\ j=1, \dots, n}} \right] \\ &= E \left[ (\vec{X} - E[\vec{X}])(\vec{X} - E[\vec{X}])^T \right] \end{aligned} \quad (1.3.22)$$

Kovarianzmatrix (Varianz-Kovarianz-Matrix) von  $\vec{X}$  und

$$R_{\vec{X}} = \left( \rho_{X_i, X_j} \right)_{\substack{i=1, \dots, n \\ j=1, \dots, n}} = \vec{\sigma}^{-1} \text{Kov}(\vec{X}) \vec{\sigma}^{-1} \quad (1.3.23)$$

Korrelationsmatrix von  $\vec{X}$ . Dabei ist  $\vec{\sigma}^{-1} = \text{diag}((\sigma_i^{-1})_{i=1, \dots, n})$  mit  $\sigma_i^{-1} = \text{Var}(X_i)^{-\frac{1}{2}}$ , falls  $\text{Var}(X_i) > 0$ , und  $\sigma_i^{-1} = 0$ , falls  $\text{Var}(X_i) = 0$ .

6. Es seien  $\vec{X}_1, \dots, \vec{X}_l$  Zufallsvektoren mit existierenden zweiten Momenten. Sie heißen paarweise unkorreliert genau dann, wenn  $\Sigma_{\vec{X}_i, \vec{X}_j} = O$  für alle  $i \neq j$ .

### Satz 1.3.6. Momente linear transformierter Zufallsmatrizen und Zufallsvektoren

1. Es seien  $\mathcal{X}$  und  $\mathcal{Y}$  ( $m \times n$ )-Zufallsmatrizen mit existierenden Erwartungswerten. Es sei  $A$  eine ( $k \times m$ )-Matrix von Konstanten und  $B$  eine ( $n \times l$ )-Matrix von Konstanten. Dann gilt

$$(a) \quad E[\mathcal{X} + \mathcal{Y}] = E[\mathcal{X}] + E[\mathcal{Y}]$$

$$(b) \quad E[A\mathcal{X}] = \left( E \left[ \sum_{\nu=1}^m a_{i\nu} X_{\nu j} \right] \right) = \left( \sum_{\nu=1}^m a_{i\nu} E[X_{\nu j}] \right) = AE[\mathcal{X}]$$

$$(c) \quad E[A\mathcal{X}B] = AE[\mathcal{X}B] = A(E[B^T \mathcal{X}^T])^T = A(B^T E[\mathcal{X}^T])^T = AE[\mathcal{X}]B$$

2. Es sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein  $n$ -dimensionaler Zufallsvektor,  $\vec{Y} = (Y_1, \dots, Y_m)^T$  ein  $m$ -dimensionaler Zufallsvektor,  $A$  eine  $(k \times n)$ -Matrix und  $B$  eine  $(l \times m)$ -Matrix.

(a) Existieren die Erwartungswerte der Komponenten von  $\vec{X}$ , dann gilt:

$$E[A\vec{X}] = AE[\vec{X}] \quad (1.3.24)$$

(b) Existieren die zweiten Momente der Komponenten von  $\vec{X}$  und  $\vec{Y}$  dann gilt:

$$\begin{aligned} \text{Kov}(A\vec{X}, B\vec{Y}) &= E[(A\vec{X} - E[A\vec{X}])(B\vec{Y} - E[B\vec{Y}])^T] \\ &= E[A(\vec{X} - E[\vec{X}])(\vec{Y} - E[\vec{Y}])^T B^T] \\ &= AE[(\vec{X} - E[\vec{X}])(\vec{Y} - E[\vec{Y}])^T] B^T \\ &= AKov(\vec{X}, \vec{Y})B^T \end{aligned} \quad (1.3.25)$$

(c) Existieren die zweiten Momente der Komponenten von  $\vec{X}$ , dann gilt:

$$\text{Kov}(A\vec{X}) = \text{Kov}(A\vec{X}, A\vec{X}) = AKov(\vec{X})A^T$$

### Satz 1.3.7. Eigenschaften der Kovarianzmatrix und Normierung von Zufallsvektoren

Es sei  $\vec{X} = (X_1, \dots, X_n)^T$  ein  $n$ -dimensionaler Zufallsvektor mit existierenden zweiten Momenten der Komponenten.

1. Es sei  $\vec{a}$  ein  $n$ -dimensionaler Vektor, dann gilt

$$\begin{aligned} \text{Var}(\vec{a}^T \vec{X}) &= \text{Kov}(\vec{a}^T \vec{X}) = \vec{a}^T \text{Kov}(\vec{X}) \vec{a} \geq 0 \\ „ = “ &\Leftrightarrow P(\vec{a}^T \vec{X} = E[\vec{a}^T \vec{X}]) = 1 \end{aligned} \quad (1.3.26)$$

2. Die Kovarianzmatrix von  $\vec{X}$  ist symmetrisch und positiv semidefinit. Sie ist genau dann positiv definit, wenn es keinen Vektor  $\vec{a} \neq \vec{0}$  und keine Konstante  $c$  derart gibt, dass  $\vec{a}^T \vec{X} = c$  mit Wahrscheinlichkeit 1 gilt. Ist der Rang von  $\text{Kov}(\vec{X})$  gleich  $k$  mit  $k < n$ , dann heißt die Verteilung von  $\vec{X}$  singulär (entartet) vom Rang  $k$ .

3. Ist die Kovarianzmatrix von  $\vec{X}$  positiv definit, dann existiert eine reguläre Matrix  $A$  mit  $\text{Kov}(\vec{X}) = AA^T$  (siehe Bemerkung 1), und es gilt:

$$(a) E[A^{-1}(\vec{X} - E[\vec{X}])] = \vec{0}$$

$$(b) \text{Kov}(A^{-1}(\vec{X} - E[\vec{X}])) = A^{-1} \text{Kov}(\vec{X})(A^{-1})^T = A^{-1}AA^T(A^{-1})^T = I$$

### Bemerkung 1.

Es sei  $C$  eine positiv-semidefinite symmetrische Matrix der Ordnung  $n$  vom Rang  $r$ . Dann existiert eine Orthonormalbasis aus Eigenvektoren  $\vec{q}_1 \dots \vec{q}_n$  zu nichtnegativen Eigenwerten  $d_1 \geq \dots \geq d_r > d_{r+1} = \dots = d_n = 0$  für  $C$ . Es sei  $Q = [\vec{q}_1 \vec{q}_2 \dots \vec{q}_n]$ ,  $Q_r = [\vec{q}_1 \vec{q}_2 \dots \vec{q}_r]$ ,  $D = \text{diag}(d_1, \dots, d_n)$ ,  $D_r = \text{diag}(d_1, \dots, d_r)$ ,  $D^{1/2} = \text{diag}(d_1^{1/2}, \dots, d_n^{1/2})$  und  $D_r^{1/2} =$

$\text{diag}(d_1^{1/2}, \dots, d_r^{1/2})$ . Dann ergibt sich mit  $A = QD^{1/2}$ ,  $A_r = Q_r D_r^{1/2}$  und  $B = QD^{1/2}Q^T$ , da  $Q$  eine Orthogonalmatrix ist und  $CQ = QD$  gilt:

$$\begin{aligned} C &= QDQ^T = QD^{1/2}(QD^{1/2})^T = AA^T \\ &= Q_r D_r Q_r^T = Q_r D_r^{1/2}(Q_r D_r^{1/2})^T = A_r A_r^T \\ &= QD^{1/2}Q^T QD^{1/2}Q^T = BB = B^2. \end{aligned} \quad (1.3.27)$$

$A_r$  ist eine  $(n \times r)$ -Matrix vom Rang  $r$ . Man bezeichnet  $B$  auch als Wurzel von  $C$  und benutzt die Schreibweise  $B = C^{1/2}$ .

**Definition 1.3.5. Multipler Korrelationskoeffizient**

Es sei  $Y$  eine Zufallsgröße und  $\vec{X}$  ein  $n$ -dimensionaler Zufallsvektor über einem gemeinsamen Wahrscheinlichkeitsraum mit existierenden zweiten Momenten. Die Größe

$$\rho_{Y;\vec{X}} = \rho_{Y;X_1,\dots,X_n} = \max_{\vec{a} \in \mathbb{R}^n \setminus \{\vec{0}\}} \rho_{Y,\vec{a}^T \vec{X}} \quad (1.3.28)$$

heißt multipler Korrelationskoeffizient zwischen  $Y$  und  $\vec{X}$ , und  $\rho_{Y;\vec{X}}^2$  heißt multiples Bestimmtheitsmaß.

## Kapitel 2

# Das lineare Modell

### 2.1 Die Methode der kleinsten Quadrate und Schätzbarkeit

#### Satz 2.1.1. Methode der kleinsten Quadrate

Es sei  $\vec{y} \in \mathbb{R}^n$  ein konkreter Wert, eine Realisation des  $n$ -dimensionalen Beobachtungsvektors  $\vec{Y}$  des linearen Modells

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\beta} \in \mathbb{R}^k \quad (2.1.1)$$

Die Minimierungsaufgabe

$$\|\vec{y} - X\hat{\beta}(\vec{y})\| = \min_{\vec{\beta} \in \mathbb{R}^k} \|\vec{y} - X\vec{\beta}\|, \quad (2.1.2)$$

ist lösbar.

$\hat{\beta}(\vec{y})$  ist Lösung dieser Minimierungsaufgabe genau dann, wenn  $\hat{\beta}(\vec{y})$  die Normalgleichungen

$$X^T X \hat{\beta}(\vec{y}) = X^T \vec{y} \quad (2.1.3)$$

erfüllt.

#### Beweis:

Die Aussagen des Satzes ergeben sich direkt aus dem Korollar 1 zu Satz 1.2.7.

#### Bemerkung 1.

Für jede Lösung  $\hat{\beta}(\vec{y})$  der Minimierungsaufgabe bzw. der Normalgleichungen ist

$$\hat{y} = X\hat{\beta}(\vec{y}) \quad (2.1.4)$$

der eindeutig bestimmte Vektor im Spaltenraum von  $X$  (Menge der aufgrund des Modells möglichen Erwartungswerte des Beobachtungsvektors  $\vec{Y}$ ) mit minimalem Abstand von  $\vec{y}$  und somit eine Invariante der Lösungen der Minimierungsaufgabe bzw. der Normalgleichungen.

$\hat{y}$  ist das eindeutig bestimmte Bild von  $\vec{y}$  unter der orthogonalen Projektion auf den Spaltenraum der Designmatrix  $X$  des linearen Modells:

$$\hat{y} = P_{\mathfrak{M}(X)} \vec{y}. \quad (2.1.5)$$

(siehe Korollar 1 zu Satz 1.2.7)

Für den minimalen Abstand des Beobachtungsvektors  $\vec{y}$  vom Spaltenraum der Designmatrix  $X$  ergibt sich aufgrund der Symmetrie und der Idempotenz der Matrix einer Orthogonalprojektion:

$$\begin{aligned} \|\vec{y} - \hat{y}\|^2 &= (\vec{y} - \hat{y})^T (\vec{y} - \hat{y}) = (\vec{y} - X\hat{\beta}(\vec{y}))^T (\vec{y} - X\hat{\beta}(\vec{y})) \\ &= (\vec{y} - P_{\mathfrak{M}(X)} \vec{y})^T (\vec{y} - P_{\mathfrak{M}(X)} \vec{y}) = \vec{y}^T (I - P_{\mathfrak{M}(X)}) \vec{y}, \end{aligned} \quad (2.1.6)$$

hierbei ist  $I - P_{\mathfrak{M}(X)}$  die Matrix der Orthogonalprojektion auf das orthogonale Komplement des Spaltenraumes von  $X$  (siehe Satz 1.2.6).

Die Vektoren  $\hat{y} = P_{\mathfrak{M}(X)} \vec{y} \in \mathfrak{M}(X)$  und  $\vec{y} - \hat{y} = (I - P_{\mathfrak{M}(X)}) \vec{y} \in \mathfrak{M}(X)^\perp$  sind orthogonal.

### Bemerkung 2.

Für jede verallgemeinerte Inverse  $(X^T X)^-$  von  $X^T X$  ist

$$\hat{\beta}(\vec{y}) = (X^T X)^- X^T \vec{y} \quad (2.1.7)$$

eine Lösung der Minimierungsaufgabe und somit der Normalgleichungen und es gilt:

$$\hat{y} = P_{\mathfrak{M}(X)} \vec{y} = X(X^T X)^- X^T \vec{y}. \quad (2.1.8)$$

(siehe Satz 1.2.10)

### Bemerkung 3.

Die Normalgleichungen sind genau dann eindeutig lösbar, wenn  $X^T X$  regulär ist, das heißt, wenn  $X$  maximalen Spaltenrang besitzt. In diesem Fall lautet die Lösung

$$\hat{\beta}(\vec{y}) = (X^T X)^{-1} X^T \vec{y}. \quad (2.1.9)$$

und die Matrix der orthogonalen Projektion auf den Spaltenraum der Designmatrix  $X$  ergibt sich zu

$$P_{\mathfrak{M}(X)} = X(X^T X)^{-1} X^T. \quad (2.1.10)$$

### Satz und Definition 2.1.2. Kleinste-Quadrate Schätzung

1. Es sei

$$\vec{\mu}_{\vec{Y}} = (\mu_{Y_1}, \dots, \mu_{Y_n})^T = E[\vec{Y}] = X\vec{\beta} \quad (2.1.11)$$

der Erwartungswert des Vektors der beobachteten Größen des durch Gleichung (2.1.1) gegebenen linearen Modells und  $\hat{\beta}(\vec{y})$  eine Lösung der Normalgleichungen (2.1.3). Die Abbildung

$$\hat{\mu}_{\vec{Y}} : \vec{y} \mapsto \hat{\mu}_{\vec{Y}}(\vec{y}) = \hat{y} = X\hat{\beta}(\vec{y}) = P_{\mathfrak{M}(X)} \vec{y} \quad (2.1.12)$$

beziehungsweise die zugehörige Zufallsgröße

$$\hat{\mu}_{\vec{Y}} = (\hat{\mu}_{Y_1}, \dots, \hat{\mu}_{Y_n})^T = \hat{\mu}_{\vec{Y}}(\vec{Y}) = \hat{Y} = X\hat{\beta} = P_{\mathfrak{M}(X)}\vec{Y} \quad (2.1.13)$$

mit  $\hat{\beta} = \hat{\beta}(\vec{Y})$  heißt *Kleinste-Quadrate-Schätzfunktion*, *Kleinste-Quadrate-Schätzung* oder *Kleinste-Quadrate-Schätzer* für  $\vec{\mu}_{\vec{Y}}$ .  $\hat{\mu}_{\vec{Y}}$  ist ein linearer erwartungstreuer Schätzer für  $\vec{\mu}_{\vec{Y}}$  mit der Kovarianzmatrix

$$\Sigma_{\hat{\mu}_{\vec{Y}}} = \Sigma_{\hat{Y}} = \sigma^2 P_{\mathfrak{M}(X)} . \quad (2.1.14)$$

$\hat{\mu}_{Y_i}$  heißt *Kleinste-Quadrate-Schätzung* für  $\mu_{Y_i}$ .

2. Die Designmatrix  $X$  des durch Gleichung (2.1.1) gegebenen linearen Modells besitze vollen Spaltenrang, dann heißt die Abbildung

$$\hat{\beta} : \vec{y} \mapsto \hat{\beta}(\vec{y}) = (X^T X)^{-1} X^T \vec{y} \quad (2.1.15)$$

beziehungsweise die zugehörige Zufallsgröße

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)^T = \hat{\beta}(\vec{Y}) = (X^T X)^{-1} X^T \vec{Y} \quad (2.1.16)$$

*Kleinste-Quadrate-Schätzfunktion*, *Kleinste-Quadrate-Schätzung* oder *Kleinste-Quadrate-Schätzer* für  $\vec{\beta}$ .  $\hat{\beta}$  ist ein linearer erwartungstreuer Schätzer für  $\vec{\beta}$  mit der Kovarianzmatrix

$$\Sigma_{\hat{\beta}} = \sigma^2 (X^T X)^{-1} . \quad (2.1.17)$$

$\hat{\beta}_i$  heißt *Kleinste-Quadrate-Schätzung* für  $\beta_i$ .

3. Für das durch Gleichung (2.1.1) gegebenen lineare Modell sei

$$\hat{\beta}_* = (X^T X)^- X^T \vec{Y} \quad (2.1.18)$$

für eine verallgemeinerte Inverse  $(X^T X)^-$  von  $X^T X$ .  $\hat{\beta}_*$  ist eine Lösung der Normalgleichungen (2.1.3) (siehe Bemerkung 2 zu Satz 2.1.1) mit

$$E[\hat{\beta}_*] = (X^T X)^- X^T X \vec{\beta} \quad (2.1.19)$$

$$\Sigma_{\hat{\beta}_*} = \sigma^2 (X^T X)^- X^T X ((X^T X)^-)^T \quad (2.1.20)$$

und somit ein linearer erwartungstreuer Schätzer für die lineare Parameterfunktion

$$\vec{\beta}_* = \vec{\psi}(\vec{\beta}) = (X^T X)^- X^T X \vec{\beta} \quad (2.1.21)$$

Durch die Abbildung  $\vec{\psi}$  wird für jeden Parametervektor  $\vec{\beta}$  bzw. jeden Erwartungswertvektor  $\vec{\mu}_{\vec{Y}}$  aus dem Spaltenraum von  $X$  ein eindeutig bestimmter Parametervektor  $\vec{\beta}_* = (X^T X)^- X^T \vec{\mu}_{\vec{Y}}$  mit

$$\vec{\mu}_{\vec{Y}} = X \vec{\beta} = X \vec{\beta}_* = X \vec{\psi}(\vec{\beta}) , \quad (2.1.22)$$

festgelegt.

Für jede Lösung  $\hat{\beta}$  der Normalgleichungen (2.1.3) gilt:

$$\hat{\beta}_* = \vec{\psi}(\hat{\beta}). \quad (2.1.23)$$

Die Schätzung  $\hat{\psi} = \vec{\psi}(\hat{\beta})$  heißt *Kleinste-Quadrate-Schätzfunktion*, *Kleinste-Quadrate-Schätzung* oder *Kleinste-Quadrate-Schätzer* für die lineare Parameterfunktion  $\vec{\psi}$ .

**Beweis:**

1. Beweis der Erwartungstreue und Bestimmung der Kovarianzmatrix von  $\hat{\mu}_{\vec{Y}}$  aus (2.1.13):

$$E[\hat{\mu}_{\vec{Y}}] = E[P_{\mathfrak{M}(X)} \vec{Y}] = P_{\mathfrak{M}(X)} E[\vec{Y}] = P_{\mathfrak{M}(X)} X \vec{\beta} = X \vec{\beta} = \vec{\mu}_{\vec{Y}} \quad (2.1.24)$$

$$\Sigma_{\hat{\mu}_{\vec{Y}}} = \Sigma_{P_{\mathfrak{M}(X)} \vec{Y}} = P_{\mathfrak{M}(X)} \Sigma_{\vec{Y}} P_{\mathfrak{M}(X)}^T = \sigma^2 P_{\mathfrak{M}(X)} \quad (2.1.25)$$

2. Beweis der Erwartungstreue und Bestimmung der Kovarianzmatrix von  $\hat{\beta}$  aus (2.1.16):

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T \vec{Y}] = (X^T X)^{-1} X^T E[\vec{Y}] = (X^T X)^{-1} X^T X \vec{\beta} = \vec{\beta} \quad (2.1.26)$$

$$\Sigma_{\hat{\beta}} = \Sigma_{(X^T X)^{-1} X^T \vec{Y}} = (X^T X)^{-1} X^T \Sigma_{\vec{Y}} ((X^T X)^{-1} X^T)^T = \sigma^2 (X^T X)^{-1} \quad (2.1.27)$$

3. Da  $(X^T X)^{-1} X^T$  nach Satz 1.2.10 eine verallgemeinerte Inverse von  $X$  ist, ergibt sich

$$X \vec{\beta}_* = X (X^T X)^{-1} X^T X \vec{\beta} = X \vec{\beta}. \quad (2.1.28)$$

Für jede Lösung  $\hat{\beta}$  der Normalgleichungen (2.1.3) gilt:

$$\hat{\psi} = \vec{\psi}(\hat{\beta}) = (X^T X)^{-1} X^T X \hat{\beta} = (X^T X)^{-1} X^T \vec{Y} = \hat{\beta}_* \quad (2.1.29)$$

**Bemerkung 1.**

Die Abbildung  $\vec{\psi}(\vec{\beta}) = (X^T X)^{-1} X^T X \vec{\beta} = \vec{\beta}_*$  ist die Projektion des  $\mathbb{R}^k$  längs des Kerns von  $X$  auf  $\mathfrak{M}(\vec{\psi})$ , ein Komplement des Kerns von  $X$ . Die Abbildung  $\vec{\psi}$  ist eine Projektion, da die zugehörige Abbildungsmatrix idempotent ist:

$$((X^T X)^{-1} X^T X)^2 = (X^T X)^{-1} X^T X (X^T X)^{-1} X^T X = (X^T X)^{-1} X^T X \quad (2.1.30)$$

da  $(X^T X)^{-1}$  eine verallgemeinerte Inverse von  $X^T X$  ist. Daher wird  $\vec{\beta}_*$  auch als projizierter Parametervektor bezeichnet.

**Bemerkung 2.**

Bei der Wahl von  $X$  sollte man - sofern nicht andere Überlegungen dagegen sprechen - so vorgehen, dass  $X^T X$  eine reguläre Diagonalmatrix ist. Dann ist auch  $(X^T X)^{-1}$  eine Diagonalmatrix, wodurch die numerische Rechnung wesentlich vereinfacht wird und sich unkorrelierte Parameterschätzungen  $\hat{\beta}_i$  ergeben.



**Bemerkung 3.**

Bei der Regressionsanalyse hat  $X$  in aller Regel vollen Spaltenrang, während bei der Varianzanalyse fast immer eine Designmatrix  $X$  mit Rangdefekt gewählt wird. Daher ist es sogar üblich, abweichend von der Definition 1.1.1, das Regressionsmodell und das Modell der Varianzanalyse über diese Eigenschaft zu definieren.

**Beispiel 2.1.1. Einfache lineare Regression**

Das durch die Gleichungen

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i, \quad E[\epsilon_i] = 0, \quad \text{Var}(\epsilon_i) = \sigma^2 > 0 \quad (i = 1, \dots, n) \\ \text{Kov}(\epsilon_i, \epsilon_j) &= 0 \quad (1 \leq i < j \leq n) \end{aligned} \quad (2.1.31)$$

gegebene einfache lineare Regressionsmodell lautet in Matrixschreibweise:

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\epsilon} = \sigma^2 I. \quad (2.1.32)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \quad (2.1.33)$$

Die Kleinste-Quadrate-Schätzung für  $\vec{\beta}$  ergibt sich wie folgt:

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad X^T \vec{Y} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n x_i Y_i \end{pmatrix} \quad (2.1.34)$$

$$\begin{aligned} (X^T X)^{-1} &= \frac{1}{n \sum_{i=1}^n x_i^2 - n^2 \bar{x}^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \\ &= \frac{1}{n \sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}, \end{aligned} \quad (2.1.35)$$

$$\text{falls } s_x^2 = s_{xx} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \neq 0$$

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T \vec{Y} = \frac{1}{n^2 s_{xx}} \begin{pmatrix} n\bar{Y} \sum_{i=1}^n x_i^2 - n\bar{x} \sum_{i=1}^n x_i Y_i \\ -n\bar{Y} n\bar{x} + n \sum_{i=1}^n x_i Y_i \end{pmatrix} \\ &= \frac{1}{n^2 s_{xx}} \begin{pmatrix} n\bar{Y} n(s_{xx} + \bar{x}^2) - n\bar{x} n(s_{xY} + \bar{x}\bar{Y}) \\ n^2 s_{xY} \end{pmatrix} = \begin{pmatrix} \bar{Y} - \frac{s_{xY}}{s_{xx}} \bar{x} \\ \frac{s_{xY}}{s_{xx}} \end{pmatrix} \end{aligned} \quad (2.1.36)$$

$$\text{mit } s_{xY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) = \frac{1}{n} \left( \sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y} \right)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{s_{xY}}{s_{xx}} = \frac{s_Y}{s_x} r_{xY} \quad (2.1.37)$$

$$\text{mit } s_Y^2 = s_{YY} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y}), \quad r_{xY} = \frac{s_{xY}}{s_x s_Y}$$

Die Schätzungen  $\hat{\beta}_0$  und  $\hat{\beta}_1$  sind erwartungstreu und besitzen die Kovarianzmatrix

$$\begin{aligned} \text{Kov}(\hat{\beta}) &= \sigma^2 (X^T X)^{-1}, \quad \text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{n} \cdot \frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{s_{xx}}, \\ \text{Var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{n} \cdot \frac{1}{s_{xx}}, \quad \text{Kov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} = -\frac{\sigma^2}{n} \cdot \frac{\bar{x}}{s_{xx}} \end{aligned} \quad (2.1.38)$$

Betrachten wir die alternative Parametrisierung

$$\begin{aligned} Y_i &= \alpha + \beta(x_i - \bar{x}) + \epsilon_i, \quad E[\epsilon_i] = 0, \quad \text{Var}(\epsilon_i) = \sigma^2 > 0 \quad (i = 1, \dots, n) \\ \text{Kov}(\epsilon_i, \epsilon_j) &= 0 \quad (1 \leq i < j \leq n) \\ \vec{Y} &= \tilde{X} \vec{\gamma} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I \end{aligned} \quad (2.1.39)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}, \quad \vec{\gamma} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \quad (2.1.40)$$

Für die Parameter der beiden Modelle gilt

$$\beta_0 = \alpha - \beta \bar{x}, \quad \beta_1 = \beta. \quad (2.1.41)$$

Die Kleinste-Quadrate-Schätzung für  $\vec{\gamma}$  ergibt sich wie folgt:

$$\tilde{X}^T \tilde{X} = \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix}, \quad \tilde{X}^T \vec{Y} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n (x_i - \bar{x}) Y_i \end{pmatrix} \quad (2.1.42)$$

$$(\tilde{X}^T \tilde{X})^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix} \quad (2.1.43)$$

$$\begin{aligned} \hat{\gamma} &= (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \vec{Y} = \begin{pmatrix} \bar{Y} \\ \frac{s_{xY}}{s_{xx}} \end{pmatrix} \\ \hat{\alpha} &= \bar{Y}, \quad \hat{\beta} = \frac{s_{xY}}{s_{xx}} = \hat{\beta}_1 \end{aligned} \quad (2.1.44)$$

Die Schätzungen  $\hat{\alpha}$  und  $\hat{\beta}$  sind erwartungstreu und besitzen die Kovarianzmatrix

$$\text{Kov}(\hat{\gamma}) = \sigma^2 (\tilde{X}^T \tilde{X})^{-1}, \quad \text{Var}(\hat{\alpha}) = \frac{\sigma^2}{n}, \quad (2.1.45)$$

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{n} \cdot \frac{1}{s_{xx}}, \quad \text{Kov}(\hat{\alpha}, \hat{\beta}) = 0$$

Als erwartungstreue Schätzung für die Regressionsgerade

$$y(x) = \beta_0 + \beta_1 x = \alpha + \beta(x - \bar{x}) \quad (2.1.46)$$

ergibt sich

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = \hat{\alpha} + \hat{\beta}(x - \bar{x}) = \bar{Y} + \frac{s_{xY}}{s_{xx}}(x - \bar{x}) \quad (2.1.47)$$

mit der Varianz

$$\begin{aligned} \text{Var}(\hat{y}(x)) &= \text{Var}(\hat{\alpha}) + (x - \bar{x})^2 \text{Var}(\hat{\beta}) \\ &= \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \sigma^2 = \frac{\sigma^2}{n} \cdot \left( 1 + \frac{(x - \bar{x})^2}{s_{xx}} \right) \end{aligned} \quad (2.1.48)$$

Die arithmetischen Mittel  $\bar{Y}$  und  $\bar{x}$  erfüllen die lineare Beziehung bei Verwendung der Kleinst-Quadrat-Schätzer exakt, der Punkt  $(\bar{x}, \bar{Y})$  liegt auf der geschätzten Regressionsgeraden.

### Beispiel 2.1.2. Einfache Varianzanalyse

Das durch die Gleichungen

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad E[\epsilon_{ij}] = 0, \quad E[Y_{ij}] = \mu_i, \quad \text{Var}(\epsilon_{ij}) = \sigma^2 > 0, \quad (2.1.49)$$

$$\text{Kov}(\epsilon_{ij}, \epsilon_{i'j'}) = 0, \quad (i, j) \neq (i', j') \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1), \quad n = \sum_{i=1}^k n_i$$

gegebene Modell der einfachen Varianzanalyse ergibt in Matrizenform mit  $\vec{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$  und  $\vec{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$  für  $i = 1, \dots, k$

$$\vec{Y} = \tilde{X}\vec{\gamma} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I. \quad (2.1.50)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_k \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \vec{1}_{n_1} & \vec{0}_{n_1} & \dots & \vec{0}_{n_1} \\ \vec{0}_{n_2} & \vec{1}_{n_2} & \dots & \vec{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{0}_{n_k} & \vec{0}_{n_k} & \dots & \vec{1}_{n_k} \end{bmatrix}, \quad \vec{\gamma} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{bmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \\ \vdots \\ \vec{\epsilon}_k \end{bmatrix} \quad (2.1.51)$$

Die Parameter  $\mu_i$  sind die Erwartungswerte bzw. (erwarteten) Mittelwerte der Stufen des zu untersuchenden Faktors, der den Stichproben zugrundeliegenden Grundgesamtheiten. Die Designmatrix  $\tilde{X}$  hat vollen Spaltenrang, daher ergibt sich die Kleinst-Quadrat-Schätzung wie folgt

$$\tilde{X}^T \tilde{X} \hat{\gamma} = \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_k \end{pmatrix} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_k \end{pmatrix} = \tilde{X}^T \vec{Y} = \begin{pmatrix} Y_{1.} \\ Y_{2.} \\ \vdots \\ Y_{k.} \end{pmatrix} \quad (2.1.52)$$

$$\hat{\mu}_i = \bar{Y}_{i.} = \frac{1}{n_i} Y_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}, \quad i = 1, \dots, k \quad (2.1.53)$$

Die erwartungstreuen Schätzungen  $\hat{\mu}_i$ , die Mittelwerte der Stichproben, der Parameter  $\mu_i$  besitzen die Kovarianzmatrix

$$\begin{aligned} \text{Kov}(\hat{\gamma}) &= \sigma^2(\tilde{X}^T \tilde{X})^{-1} = \text{diag}\left(\frac{\sigma^2}{n_1}, \dots, \frac{\sigma^2}{n_k}\right), \\ \text{Var}(\hat{\mu}_i) &= \frac{\sigma^2}{n_i}, \quad \text{Kov}(\hat{\mu}_i, \hat{\mu}_j) = 0, \quad i \neq j \quad (i, j = 1, \dots, k) \end{aligned} \quad (2.1.54)$$

Betrachten wir die alternative Parametrisierung

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + \epsilon_{ij}, \quad E[\epsilon_{ij}] = 0, \quad E[Y_{ij}] = \mu_i = \mu + \alpha_i, \quad \text{Var}(\epsilon_{ij}) = \sigma^2 > 0, \\ \text{Kov}(\epsilon_{ij}, \epsilon_{i'j'}) &= 0, \quad (i, j) \neq (i', j') \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1), \quad n = \sum_{i=1}^k n_i \\ \vec{Y} &= X\vec{\beta} + \vec{\epsilon} \quad E[\vec{\epsilon}] = \vec{0} \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I. \end{aligned} \quad (2.1.55)$$

wobei die  $\mu_i$  die Parameter des ursprünglichen Modells sind. Den Parameter  $\mu$  bezeichnen wir als das Gesamtmittel und  $\alpha_i$  als den Effekt der  $i$ -ten Stufe des zu untersuchenden Faktors. Die Matrix und die Vektoren sind gegeben durch

$$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_k \end{bmatrix}, \quad X = \begin{bmatrix} \vec{1}_{n_1} & \vec{0}_{n_1} & \vec{0}_{n_1} & \dots & \vec{0}_{n_1} \\ \vec{1}_{n_2} & \vec{0}_{n_2} & \vec{1}_{n_2} & \dots & \vec{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{1}_{n_k} & \vec{0}_{n_k} & \vec{0}_{n_k} & \dots & \vec{1}_{n_k} \end{bmatrix}, \quad \vec{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{bmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \\ \vdots \\ \vec{\epsilon}_k \end{bmatrix} \quad (2.1.56)$$

Die Normalgleichungen ergeben sich zu

$$X^T X \hat{\beta} = \begin{pmatrix} n & n_1 & n_2 & \dots & n_k \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_k & 0 & 0 & \dots & n_k \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_k \end{pmatrix} = X^T \vec{Y} = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ \vdots \\ Y_{k.} \end{pmatrix} \quad (2.1.57)$$

mit  $Y_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}$  und  $\bar{Y}_{..} = \frac{1}{n} Y_{..}$ . Der Rang der  $((k+1) \times (k+1))$ -Matrix  $X^T X$  ist  $k$ , das heißt, die Matrix ist singulär.

Eine verallgemeinerte Inverse der Matrix  $X^T X$  ist (siehe Übung)

$$(X^T X)^- = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{n_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{n_k} \end{pmatrix}. \quad (2.1.58)$$

Mit dieser ergibt sich als Lösung der Normalgleichungen

$$\hat{\beta}_*^{(1)} = (X^T X)^- X^T \vec{Y} = (0, \bar{Y}_{1.}, \dots, \bar{Y}_{k.})^T. \quad (2.1.59)$$

Dieses ist eine erwartungstreue Schätzung der linearen Parameterfunktion

$$\vec{\beta}_*^{(1)} = \vec{\psi}_1(\vec{\beta}) = (X^T X)^{-1} X^T X \vec{\beta} = E[\hat{\beta}_*^{(1)}] = (0, \mu + \alpha_1, \dots, \mu + \alpha_k)^T. \quad (2.1.60)$$

Die Hermite-Form der Matrix  $X^T X$ , die sich als Produkt einer verallgemeinerten Inversen von  $X^T X$  und der Matrix  $X^T X$  darstellen lässt, lautet (siehe Übung):

$$H = (X^T X)^{-1} X^T X = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & -1 \\ 0 & 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (2.1.61)$$

sie führt auf die lineare Parameterfunktion

$$\vec{\beta}_*^{(2)} = \vec{\psi}_2(\vec{\beta}) = (X^T X)^{-1} X^T X \vec{\beta} = (\mu + \alpha_k, \alpha_1 - \alpha_k, \dots, \alpha_{k-1} - \alpha_k, 0)^T, \quad (2.1.62)$$

die durch die folgende Lösung der Normalgleichungen

$$\hat{\beta}_*^{(2)} = (X^T X)^{-1} X^T \vec{Y} = \psi_2(\hat{\beta}) = (\bar{Y}_k, \bar{Y}_1 - \bar{Y}_k, \dots, \bar{Y}_{k-1} - \bar{Y}_k, 0)^T \quad (2.1.63)$$

erwartungstreu geschätzt wird. Hierbei ist  $\hat{\beta}$  eine beliebige Lösung der Normalgleichungen, zum Beispiel  $\hat{\beta}_*^{(1)}$ .

### Definition 2.1.1. Schätzbare lineare Parameterfunktionen

Eine ( $q$ -dimensionale) lineare Funktion  $\vec{\psi}(\vec{\beta}) = (\psi_1(\vec{\beta}), \dots, \psi_q(\vec{\beta}))^T = C\vec{\beta}$  ( $C \in \mathbb{R}^{q \times k}$ ) des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells heißt schätzbar genau dann, wenn eine in  $\vec{Y}$  lineare erwartungstreue Schätzung für  $\vec{\psi}$  existiert, wenn es also eine Funktion  $\tilde{\psi} = (\tilde{\psi}_1, \dots, \tilde{\psi}_q) = A\vec{Y}$  ( $A \in \mathbb{R}^{q \times n}$ ) mit

$$E[\tilde{\psi}] = E[A\vec{Y}] = \vec{\psi}(\vec{\beta}) = C\vec{\beta} \quad \forall \vec{\beta} \in \mathbb{R}^k \quad (2.1.64)$$

gibt.

### Bemerkung 1.

Eine ( $q$ -dimensionale) lineare Funktion  $\vec{\psi}(\vec{\beta}) = (\psi_1(\vec{\beta}), \dots, \psi_q(\vec{\beta}))^T = C\vec{\beta}$  des Parametervektors  $\vec{\beta}$  ist schätzbar genau dann, wenn die  $q$  Komponentenfunktionen  $\psi_1, \dots, \psi_q$  (eindimensionale) schätzbare Funktionen des Parametervektors sind.

### Satz 2.1.3.

Der Parametervektor  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells ist schätzbar genau dann, wenn die Designmatrix  $X$  vollen Spaltenrang ( $\text{rg}(X) = k$ ) hat.

**Beweis:**

Dass die Bedingung  $rg(X) = k$  hinreichend für die Schätzbarkeit von  $\vec{\beta}$  ist, folgt aus Satz 2.1.2. Nachzuweisen ist die Notwendigkeit dieser Bedingung. Es sei  $\tilde{\beta} = A\vec{Y}$  ein linearer erwartungstreuer Schätzer für  $\vec{\beta}$ . Dann ist  $A$  eine  $k \times n$ -Matrix mit

$$\vec{\beta} = E[\tilde{\beta}] = E[A\vec{Y}] = AE[\vec{Y}] = AX\vec{\beta} \quad \forall \vec{\beta} \in \mathbb{R}^k. \quad (2.1.65)$$

Daraus folgt  $AX = I_k$ . Mit (siehe Gleichung (1.2.45))  $k = rg(I_k) = rg(AX) \leq rg(X) \leq k$  ergibt sich  $rg(X) = k$ .

**Satz 2.1.4. Schätzbare lineare Parameterfunktionen**

Eine (eindimensionale) lineare Funktion  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  ( $\vec{c} \in \mathbb{R}^k$ ) des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells ist schätzbar genau dann, wenn es einen Vektor  $\vec{a}$  ( $\vec{a} \in \mathbb{R}^n$ ) gibt mit

$$\vec{c}^T = \vec{a}^T X \quad (2.1.66)$$

das heißt genau dann, wenn  $\vec{c}$  im Zeilenraum der Designmatrix  $X$  liegt. Die Abbildung, die jeder schätzbaren Funktion  $\vec{c}^T \vec{\beta}$  den Vektor  $\vec{c}$  zuordnet, ist ein Isomorphismus zwischen dem Vektorraum der schätzbaren linearen Funktionen und dem Zeilenraum von  $X$ . Die eindimensionalen linearen schätzbaren Funktionen bilden somit einen Vektorraum der Dimension  $rg(X)$ .

**Beweis:**

Es sei  $\vec{c}^T \vec{\beta}$  ( $\vec{c} \in \mathbb{R}^k$ ) schätzbar. Dann gibt es einen erwartungstreuen linearen Schätzer  $\vec{a}^T \vec{Y}$  ( $\vec{a} \in \mathbb{R}^n$ ) für  $\vec{c}^T \vec{\beta}$ , das heißt

$$\vec{c}^T \vec{\beta} = E[\vec{a}^T \vec{Y}] = \vec{a}^T E[\vec{Y}] = \vec{a}^T X \vec{\beta} \quad \forall \vec{\beta} \in \mathbb{R}^k. \quad (2.1.67)$$

Hieraus folgt  $\vec{c}^T = \vec{a}^T X$ .

Gilt umgekehrt  $\vec{c}^T = \vec{a}^T X$ , so folgt

$$E[\vec{a}^T \vec{Y}] = \vec{a}^T E[\vec{Y}] = \vec{a}^T X \vec{\beta} = \vec{c}^T \vec{\beta}, \quad (2.1.68)$$

somit ist  $\vec{a}^T \vec{Y}$  eine in  $\vec{Y}$  lineare erwartungstreue Schätzung für  $\vec{c}^T \vec{\beta}$ .

**Korollar 1.**

Die Linearkombination der Beobachtungen des durch Gleichung (2.1.1) gegebenen linearen Modells  $\vec{a}^T \vec{Y}$  ( $\vec{a} \in \mathbb{R}^n$ ) ist eine erwartungstreue Schätzung der linearen Funktion  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  genau dann, wenn

$$\vec{c}^T = \vec{a}^T X. \quad (2.1.69)$$

**Korollar 2.**

1. Eine  $q$ -dimensionale lineare Funktion  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  ( $C \in \mathbb{R}^{q \times k}$ ) des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells ist schätzbar genau dann, wenn es eine Matrix  $A$  ( $A \in \mathbb{R}^{q \times n}$ ) gibt mit

$$C = AX . \quad (2.1.70)$$

2. Eine  $q$ -dimensionale lineare Funktion  $A\vec{Y}$  ( $A \in \mathbb{R}^{q \times n}$ ) des Beobachtungsvektors  $\vec{Y}$  des durch Gleichung (2.1.1) gegebenen linearen Modells) ist eine erwartungstreue Schätzung der linearen Funktion  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  genau dann, wenn

$$C = AX . \quad (2.1.71)$$

**Korollar 3.**

Die Erwartungswerte der Beobachtungen  $Y_i$  des durch Gleichung (2.1.1) gegebenen linearen Modells mit der Designmatrix  $X = (x_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,k}}$

$$E[Y_i] = \sum_{j=1}^k x_{ij}\beta_j , \quad i = 1, \dots, n \quad (2.1.72)$$

bzw. in Matrizenform

$$E[\vec{Y}] = X\vec{\beta} \quad (2.1.73)$$

sind schätzbare lineare Funktionen, und jede schätzbare lineare Funktion  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  ist eine Linearkombination dieser Erwartungswerte:

$$\exists \vec{a} \in \mathbb{R}^n : \quad \psi(\vec{\beta}) = \vec{c}^T \vec{\beta} = \vec{a}^T X\vec{\beta} = \vec{a}^T E[\vec{Y}] \quad (2.1.74)$$

**Korollar 4.**

Es sei  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  mit  $\vec{c} \in \mathbb{R}^k$  eine (eindimensionale) schätzbare Funktion. Dann gibt es genau einen linearen erwartungstreuen Schätzer  $\vec{a}_*^T \vec{Y}$  von  $\psi$  mit  $\vec{a}_* \in \mathfrak{Im}(X)$ . Für einen beliebigen linearen erwartungstreuen Schätzer  $\tilde{\psi} = \vec{a}^T \vec{Y}$  ist  $\vec{a}_*$  die orthogonale Projektion von  $\vec{a}$  auf  $\mathfrak{Im}(X)$ .

**Beweis:**

Da  $\psi$  schätzbar ist, gibt es einen linearen erwartungstreuen Schätzer  $\tilde{\psi} = \vec{a}^T \vec{Y}$ . Es sei

$$\vec{a}_* = P_{\mathfrak{Im}(X)} \vec{a} \in \mathfrak{Im}(X) , \quad \vec{b} = \vec{a} - \vec{a}_* = (I - P_{\mathfrak{Im}(X)}) \vec{a} \in \mathfrak{Im}(X)^\perp \quad (2.1.75)$$

Dann ist

$$\begin{aligned} \psi(\vec{\beta}) &= E[\vec{a}^T \vec{Y}] = E[(\vec{a}_* + \vec{b})^T \vec{Y}] = E[\vec{a}_*^T \vec{Y}] + E[\vec{b}^T \vec{Y}] \\ &= E[\vec{a}_*^T \vec{Y}] + \vec{b}^T X\vec{\beta} = E[\vec{a}_*^T \vec{Y}] . \end{aligned} \quad (2.1.76)$$

Somit ist  $\vec{a}_*^T \vec{Y}$  ein linearer erwartungstreuer Schätzer für  $\psi$ .

Es sei  $\vec{a}_0 \in \mathfrak{M}(X)$  mit

$$E[\vec{a}_0^T \vec{Y}] = \psi(\vec{\beta}) \quad \forall \vec{\beta} \in \mathbb{R}^k \quad (2.1.77)$$

dann ergibt sich

$$0 = E[\vec{a}_0^T \vec{Y}] - E[\vec{a}_*^T \vec{Y}] = (\vec{a}_0 - \vec{a}_*)^T X \vec{\beta} \quad \forall \vec{\beta} \in \mathbb{R}^k \quad (2.1.78)$$

hieraus folgt  $\vec{a}_0 - \vec{a}_* \in \mathfrak{M}(X)^\perp$ . Da außerdem  $\vec{a}_0 - \vec{a}_* \in \mathfrak{M}(X)$  ist, muss  $\vec{a}_0 - \vec{a}_* = \vec{0}$  und somit  $\vec{a}_0 = \vec{a}_*$  gelten.

### Satz 2.1.5. Satz von Gauß-Markoff

Für jede (eindimensionale) schätzbare Funktion  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  ( $\vec{c} \in \mathbb{R}^k$ ) des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells gibt es eine eindeutig bestimmte lineare erwartungstreue Schätzung  $\hat{\psi}$  mit minimaler Varianz in der Menge aller linearen erwartungstreuen Schätzungen für  $\psi$ , und es gilt  $\hat{\psi} = \vec{c}^T \hat{\beta}$  für jede Lösung  $\hat{\beta}$  der Normalgleichungen (2.1.3). ( $\hat{\psi}$  wird auch als Kleinste-Quadrate-Schätzer für  $\psi$  bezeichnet.)

#### Beweis:

Für einen beliebigen linearen Schätzer gilt:

$$\text{Var}(\vec{a}^T \vec{Y}) = \vec{a}^T \Sigma_{\vec{Y}} \vec{a} = \sigma^2 \vec{a}^T I \vec{a} = \sigma^2 \|\vec{a}\|^2 \quad (2.1.79)$$

Es sei  $\tilde{\psi} = \vec{a}^T \vec{Y}$  eine lineare erwartungstreue Schätzung für  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  und  $\hat{\psi} = \vec{a}_*^T \vec{Y}$  die nach Korollar 4 zu Satz 2.1.4 eindeutig bestimmte erwartungstreue Schätzung von  $\psi$  mit  $\vec{a}_* \in \mathfrak{M}(X)$ . Da  $\vec{a}_* = P_{\mathfrak{M}(X)} \vec{a}$ , sind  $\vec{a}_*$  und  $\vec{a} - \vec{a}_*$  orthogonal, und es ergibt sich

$$\begin{aligned} \text{Var}(\tilde{\psi}) &= \sigma^2 \|\vec{a}\|^2 = \sigma^2 (\|\vec{a} - \vec{a}_*\|^2 + \|\vec{a}_*\|^2) \\ &\geq \sigma^2 \|\vec{a}_*\|^2 = \text{Var}(\hat{\psi}) \\ „ = “ &\Leftrightarrow \vec{a}_* = \vec{a} \Leftrightarrow \tilde{\psi} = \hat{\psi} \end{aligned} \quad (2.1.80)$$

Somit ist  $\hat{\psi}$  die eindeutig bestimmte lineare erwartungstreue Schätzung mit minimaler Varianz in der Menge aller linearen erwartungstreuen Schätzungen für  $\psi$ .

Aus  $\vec{a}_* \in \mathfrak{M}(X)$  folgt mit Gleichung (1.2.26)  $\vec{a}_* = P_{\mathfrak{M}(X)} \vec{a}_*$ , somit ergibt sich

$$\hat{\psi} = \vec{a}_*^T \vec{y} = \vec{a}_*^T P_{\mathfrak{M}(X)} \vec{y} = \vec{a}_*^T \hat{y} = \vec{a}_*^T X \hat{\beta} = \vec{c}^T \hat{\beta} = \vec{\psi}(\hat{\beta}) \quad (2.1.81)$$

für jede Lösung  $\hat{\beta}$  der Normalgleichungen.  $\vec{a}_*^T X = \vec{c}^T$  folgt aus dem Korollar 1 zu Satz 2.1.4.

#### Korollar 1.

Die Linearkombination der Beobachtungen des durch Gleichung (2.1.1) gegebenen linearen Modells  $\vec{a}^T \vec{Y}$  ( $\vec{a} \in \mathbb{R}^n$ ) ist die eindeutig bestimmte lineare erwartungstreue Schätzung mit minimaler Varianz in der Menge aller linearen erwartungstreuen Schätzungen für  $\psi(\vec{\beta}) = E[\vec{a}^T \vec{Y}]$  genau dann, wenn  $\vec{a} \in \mathfrak{M}(X)$ .



**Definition 2.1.2. Bester linearer erwartungstreuer Schätzer (BLE-Schätzer)**

Es sei  $\vec{\psi}(\vec{\beta}) = (\psi_1(\vec{\beta}), \dots, \psi_q(\vec{\beta}))^T = C\vec{\beta}$  ( $C \in \mathbb{R}^{q \times k}$ ) eine ( $q$ -dimensionale) schätzbare Funktion des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells. Der Vektor  $\hat{\psi} = (\hat{\psi}_1, \dots, \hat{\psi}_q)^T$  der nach dem Satz von Gauß-Markoff existierenden und eindeutig bestimmten linearen erwartungstreuen Schätzungen der Komponentenfunktionen  $\psi_1, \dots, \psi_q$  mit minimaler Varianz heißt bester linearer erwartungstreuer Schätzer (BLE-Schätzer) für  $\vec{\psi}$ .

**Satz 2.1.6.**

Es sei  $\hat{\psi} = A\vec{Y}$  der BLE-Schätzer für die  $q$ -dimensionale schätzbare Funktion  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  ( $C \in \mathbb{R}^{q \times k}$ ) des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells. Dann gilt  $\hat{\psi} = C\hat{\beta}$  für jede Lösung  $\hat{\beta}$  der Normalgleichungen (2.1.3), und es ergibt sich

$$E[\hat{\psi}] = E[C\hat{\beta}] = C\vec{\beta}, \quad \Sigma_{\hat{\psi}} = \sigma^2 AA^T = \sigma^2 C(X^T X)^{-1} C^T \quad (2.1.82)$$

sowie  $rg(C) = rg(A)$ .

**Beweis:**

Da  $A\vec{Y}$  eine erwartungstreue Schätzung für  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  ist, gilt  $C = AX$  nach Korollar 2 zu Satz 2.1.4. Somit ergibt sich mit  $\hat{\psi} = C\hat{\beta} = AX\hat{\beta} = A\hat{Y}$  und Gleichung (2.1.14)

$$\begin{aligned} \Sigma_{\hat{\psi}} &= \Sigma_{A\hat{Y}} = A\Sigma_{\hat{Y}}A^T = \sigma^2 AA^T \\ &= \Sigma_{A\hat{Y}} = A\Sigma_{\hat{Y}}A^T = \sigma^2 AP_{\mathcal{M}(X)}A^T = \sigma^2 AX(X^T X)^{-1}X^T A^T \\ &= \sigma^2 C(X^T X)^{-1}C^T \end{aligned} \quad (2.1.83)$$

sowie mit den Gleichungen (1.2.45) und (1.2.46)

$$rg(C) = rg(AX) \leq rg(A) = rg(AA^T) = rg(C(X^T X)^{-1}C^T) \leq rg(C). \quad (2.1.84)$$

**Korollar 1.**

Es sei  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  ( $C \in \mathbb{R}^{q \times k}$ ) eine  $q$ -dimensionale schätzbare lineare Funktion des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells mit dem BLE-Schätzer  $\hat{\psi}$  und  $D$  eine  $(p \times q)$ -Matrix, dann ist auch die Funktion  $\vec{\phi}(\vec{\beta}) = D\vec{\psi}(\vec{\beta}) = DC\vec{\beta}$  schätzbar mit dem BLE-Schätzer  $\hat{\phi} = D\hat{\psi}$ .

**Bemerkung 1.**

Es seien

$$X^T X \hat{\beta} = X^T \vec{Y} \quad (2.1.85)$$

die Normalgleichungen des durch Gleichung (2.1.1) gegebenen linearen Modells mit  $rg(X) = p$  ( $p \leq k$ ).

Nach Korollar 2 zu Satz 2.1.4 ist die lineare Funktion

$$\vec{\psi}(\vec{\beta}) = (\psi_1(\vec{\beta}), \dots, \psi_k(\vec{\beta}))^T = X^T X \vec{\beta} \quad (2.1.86)$$

schätzbar, und für jede Lösung  $\hat{\beta}$  der Normalgleichungen ist nach Satz 2.1.6

$$\hat{\psi} = X^T X \hat{\beta} = X^T \vec{Y} \quad (2.1.87)$$

der BLE-Schätzer von  $\vec{\psi}$ .

Da nach Gleichung (1.2.47)

$$\mathfrak{I}\mathfrak{m}(X^T) = \mathfrak{I}\mathfrak{m}(X^T X) = \mathfrak{I}\mathfrak{m}((X^T X)^T) \quad (2.1.88)$$

stimmen die Zeilenräume von  $X$  und  $X^T X$  überein. Somit bilden die Zeilen von  $X^T X$  ein Erzeugendensystem des Zeilenraumes von  $X$  und die Komponenten von  $\vec{\psi}$  ein Erzeugendensystem der schätzbaren linearen Funktionen. Für jede schätzbare (eindimensionale) Funktion  $\phi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  ( $\vec{c} \in \mathbb{R}^k$ ) des Parametervektors  $\vec{\beta}$  gibt es somit einen Vektor  $\vec{l} \in \mathbb{R}^k$  mit

$$\phi(\vec{\beta}) = \vec{l}^T \vec{\psi}(\vec{\beta}) = \vec{l}^T X^T X \vec{\beta} \quad (2.1.89)$$

so dass nach Korollar 1 zu Satz 2.1.6

$$\hat{\phi} = \vec{l}^T \hat{\psi} = \vec{l}^T X^T \vec{Y} \quad (2.1.90)$$

der BLE-Schätzer für  $\phi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  ist.

Ist  $\vec{\psi}_*(\vec{\beta}) = (\psi_1^*(\vec{\beta}), \dots, \psi_p^*(\vec{\beta}))$  eine  $p$ -dimensionale Funktion, die aus linear unabhängigen Komponenten von  $\vec{\psi}(\vec{\beta})$  besteht, mit dem BLE-Schätzer  $\hat{\psi}_*$ , dann gibt es einen Vektor  $\vec{l}_* \in \mathbb{R}^p$  mit

$$\phi(\vec{\beta}) = \vec{l}_*^T \vec{\psi}_*(\vec{\beta}) \quad (2.1.91)$$

$$\hat{\phi} = \vec{l}_*^T \hat{\psi}_* . \quad (2.1.92)$$

### Beispiel 2.1.3. Einfache Varianzanalyse

Für das durch die Gleichungen

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} , \quad E[\epsilon_{ij}] = 0 , \quad E[Y_{ij}] = \mu_i = \mu + \alpha_i , \quad \text{Var}(\epsilon_{ij}) = \sigma^2 > 0 ,$$

$$\text{Kov}(\epsilon_{ij}, \epsilon_{i'j'}) = 0 , \quad (i, j) \neq (i', j') \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1) , \quad n = \sum_{i=1}^k n_i$$

gegebene Modell der einfachen Varianzanalyse lauten die Normalgleichungen (siehe Beispiel 2.1.2)

$$X^T X \hat{\beta} = \begin{pmatrix} n & n_1 & n_2 & \dots & n_k \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_k & 0 & 0 & \dots & n_k \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_k \end{pmatrix} = X^T \vec{Y} = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ \vdots \\ Y_{k.} \end{pmatrix} \quad (2.1.93)$$

mit  $Y_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}$  und  $\bar{Y}_{..} = \frac{1}{n} Y_{..}$

Die erste Zeile der Matrix  $X^T X$  ist die Summe der übrigen  $k$  Zeilen, die linear unabhängig sind. Somit bilden nach Bemerkung 1 zu Satz 2.1.6 diese unabhängigen Zeilenvektoren eine Basis des Zeilenraums von  $X$  und die zugehörigen schätzbaren linearen Funktionen der Parameter

$$\psi_i^*(\vec{\beta}) = n_i \mu + n_i \alpha_i, \quad i = 1, \dots, k \quad (2.1.94)$$

mit den BLE-Schätzern

$$\hat{\psi}_i^* = n_i \hat{\mu} + n_i \hat{\alpha}_i = Y_{i\cdot}, \quad i = 1, \dots, k \quad (2.1.95)$$

eine Basis der schätzbaren linearen Funktionen. Mit Korollar 1 zu Satz 2.1.6 folgt, dass auch die Funktionen

$$\psi_i(\vec{\beta}) = \frac{1}{n_i} \psi_i^*(\vec{\beta}) = \mu + \alpha_i, \quad i = 1, \dots, k \quad (2.1.96)$$

schätzbar mit den BLE-Schätzern

$$\hat{\psi}_i = \frac{1}{n_i} \hat{\psi}_i^* = \hat{\mu} + \hat{\alpha}_i = \bar{Y}_{i\cdot}, \quad i = 1, \dots, k \quad (2.1.97)$$

sind und eine Basis der schätzbaren linearen Funktionen bilden. Hierbei bilden  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_k$  eine beliebige Lösung der Normalgleichungen.

Mit  $\vec{\psi} = (\psi_1, \dots, \psi_k)^T$  und  $\hat{\psi} = (\hat{\psi}_1, \dots, \hat{\psi}_k)^T$  lautet somit die allgemeine Form einer schätzbaren Funktion

$$\phi_{\vec{l}}(\vec{\beta}) = \vec{l}^T \vec{\psi}(\vec{\beta}) = \sum_{i=1}^k l_i (\mu + \alpha_i) = (l_1 + \dots + l_k) \mu + \sum_{i=1}^k l_i \alpha_i \quad (2.1.98)$$

mit dem BLE-Schätzer

$$\hat{\phi}_{\vec{l}} = \vec{l}^T \hat{\psi} = \sum_{i=1}^k l_i \bar{Y}_{i\cdot}. \quad (2.1.99)$$

Spezielle schätzbare Funktionen und ihre BLE-Schätzer sind somit

- $\phi_{(1,0,\dots,0)^T}(\vec{\beta}) = \psi_1(\vec{\beta}) = \mu + \alpha_1 = \mu_1$   
 $\hat{\phi}_{(1,0,\dots,0)^T} = \hat{\psi}_1 = \hat{\mu} + \hat{\alpha}_1 = \bar{Y}_{1\cdot} = \hat{\mu}_1$
- $\phi_{(1,-1,\dots,0)^T}(\vec{\beta}) = \psi_1(\vec{\beta}) - \psi_2(\vec{\beta}) = \alpha_1 - \alpha_2 = \mu_1 - \mu_2$   
 $\hat{\phi}_{(1,-1,\dots,0)^T} = \hat{\psi}_1 - \hat{\psi}_2 = \hat{\alpha}_1 - \hat{\alpha}_2 = \bar{Y}_{1\cdot} - \bar{Y}_{2\cdot} = \hat{\mu}_1 - \hat{\mu}_2$

Eine lineare schätzbare Funktion  $\phi$  der Parameter ist eine Funktion ausschließlich von  $\alpha_1, \dots, \alpha_k$ , der Effekte der Stufen des Einflussfaktors, genau dann, wenn in der durch Gleichung (2.1.98) gegebenen Darstellung von  $\phi$  gilt  $\sum_{i=1}^k l_i = 0$ . Eine lineare Funktion  $\phi_{\alpha}(\alpha_1, \dots, \alpha_k) = \sum_{i=1}^k c_i \alpha_i$  mit  $\sum_{i=1}^k c_i = 0$  heißt (linearer) Kontrast der Parameter  $\alpha_i$ . Somit ist eine lineare Funktion  $\phi_{\alpha}$  der Parameter  $\alpha_i$  genau dann schätzbar, wenn  $\phi_{\alpha}$  ein Kontrast ist. Die Kontraste der Parameter  $\alpha_i$  stimmen überein mit den Kontrasten der Parameter  $\mu_i$  des ursprünglichen Modells:  $\phi_{\alpha} = \sum_{i=1}^k c_i \alpha_i = \sum_{i=1}^k c_i \mu_i$ . Der BLE-Schätzer des Kontrasts  $\phi_{\alpha}(\alpha_1, \dots, \alpha_k) = \sum_{i=1}^k c_i \alpha_i$  ist  $\hat{\phi} = \sum_{i=1}^k c_i \bar{Y}_{i\cdot}$ .

**Definition 2.1.3. Kontraste**

Eine lineare Funktion  $\psi = \vec{c}^T \vec{\beta}$  ( $\vec{c} \in \mathbb{R}^k$ ) des Parametervektors  $\vec{\beta}$  des durch Gleichung (2.1.1) gegebenen linearen Modells mit  $\sum_{i=1}^k c_i = 0$  heißt (linearer) Kontrast.

**Bemerkung 1. Eindeutige Parametrisierung durch Nebenbedingungen bei Rangdefekt**

Der Praktiker will häufig den gesamten Koeffizientenvektor  $\vec{\beta}$  schätzen, dessen Komponenten eine von ihm festgelegte konkrete Bedeutung haben. Dieses setzt voraus, dass ein  $\vec{\beta}$ -Wert, der die Gleichung

$$X\vec{\beta} = \vec{\mu}_{\vec{y}} \quad (2.1.100)$$

erfüllt, für jeden Vektor  $\vec{\mu}_{\vec{y}} \in \mathfrak{Im}(X)$  ausgezeichnet wird. Dieses kann auf zwei Arten geschehen.

1. Der Parameterbereich wird so eingeschränkt, dass der ausgezeichnete Wert der einzige im Parameterbereich ist, der die Gleichung erfüllt. Dieses geschieht im Allgemeinen, indem ein Komplement  $\mathcal{L}$  des Kerns der Designmatrix als Parameterraum ausgewählt wird (s. Bem. 1 (1) zu Satz 1.2.8), das durch Gleichungen der Form  $H\vec{\beta} = \vec{0}$  festgelegt wird, das heißt

$$\mathcal{L} = \mathcal{L}_H = \{\vec{\beta} \mid H\vec{\beta} = \vec{0}, \vec{\beta} \in \mathbb{R}^k\} = \mathfrak{K}(H) \quad (2.1.101)$$

$H$  ist somit so zu wählen, dass

$$\{\vec{\mu} \mid \vec{\mu} = X\vec{\beta}, \vec{\beta} \in \mathbb{R}^k\} = \{\vec{\mu} \mid \vec{\mu} = X\vec{\beta}, \vec{\beta} \in \mathcal{L}_H\} \quad (2.1.102)$$

sowie

$$\forall \vec{\beta}_1, \vec{\beta}_2 \in \mathbb{R}^k : \quad X\vec{\beta}_1 = X\vec{\beta}_2, H\vec{\beta}_1 = H\vec{\beta}_2 = \vec{0} \quad \Rightarrow \quad \vec{\beta}_1 = \vec{\beta}_2 \quad (2.1.103)$$

und die Komponenten des ausgezeichneten Parameterwertes die gewünschte Bedeutung haben.

Sind diese Bedingungen erfüllt und ist  $\hat{y} = P_{\mathfrak{Im}(X)}\vec{y}$ , dann sind die Gleichungen

$$X\hat{\beta} = \hat{y} \quad (2.1.104)$$

und, da die Lösungen dieser Gleichungen mit den Lösungen der Normalgleichungen übereinstimmen (siehe Korollar 1 zu Satz 1.2.7), auch die Normalgleichungen in  $\mathcal{L}$ , d.h. unter den Nebenbedingungen  $H\vec{\beta} = \vec{0}$ , eindeutig lösbar.

2. Man bestimmt eine lineare Funktion  $\vec{\psi}$ , die jedem Parametervektor  $\vec{\beta} \in \mathbb{R}^k$  den ausgezeichneten Parameterwert zuordnet:

$$\forall \vec{\beta} \in \mathbb{R}^k : \quad \vec{\psi}(\vec{\beta}) = \vec{\beta}_* \quad \Leftrightarrow \quad X\vec{\beta} = X\vec{\beta}_*, H\vec{\beta}_* = \vec{0} \quad (2.1.105)$$

Eine Funktion  $\vec{\psi}$  mit diesen Eigenschaften existiert genau dann, wenn das Gleichungssystem in Gleichung (2.1.105) eindeutig lösbar ist. In diesem Fall ist  $\vec{\psi}$ , wie im Folgenden gezeigt

wird, eine schätzbare lineare Funktion der Parameter,  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$ . Für jede Lösung  $\hat{\beta}$  der Normalgleichungen ist dann

$$\hat{\beta}_* = \hat{\psi} = C\hat{\beta} = \vec{\psi}(\hat{\beta}) \quad (2.1.106)$$

der BLE-Schätzer von  $\vec{\psi}$  bzw.  $\vec{\beta}_*$ . Somit ist  $\hat{\beta}_*$  die eindeutig bestimmte Lösung des Gleichungssystems

$$\hat{y} = X\hat{\beta} = X\hat{\beta}_*, \quad H\hat{\beta}_* = \vec{0} \quad (2.1.107)$$

und somit auch die eindeutig bestimmte Lösung der Normalgleichungen, die die Nebenbedingungen  $H\hat{\beta} = \vec{0}$  erfüllt.

Das Gleichungssystem in Gleichung (2.1.105) ist für alle  $\vec{\beta} \in \mathbb{R}^k$  eindeutig lösbar genau dann, wenn die in den Gleichungen (2.1.102) und (2.1.103) formulierten Bedingungen erfüllt sind. In diesem Fall führen beide Ansätze zu demselben Parameterschätzwert. (Die nach dem zweiten Ansatz bestimmte Funktion  $\vec{\psi}$  ist die Projektion des  $\mathbb{R}^k$  auf  $\mathcal{L}_H$  längs  $\mathfrak{R}(X)$ .)

### Bemerkung 2.

Es sei  $X$  eine  $(n \times k)$ -Matrix mit  $\text{rg}(X) = p$  ( $p < k$ ),  $H$  eine  $(m \times k)$ -Matrix und

$$G = \begin{bmatrix} X \\ H \end{bmatrix}. \quad (2.1.108)$$

Dann ist das Gleichungssystem

$$X\vec{\beta} = X\vec{\beta}_*, \quad H\vec{\beta}_* = \vec{0} \quad (2.1.109)$$

für alle  $\vec{\beta} \in \mathbb{R}^k$  eindeutig lösbar genau dann, wenn die Zeilenräume der Matrizen  $X$  und  $H$  linear unabhängig sind, das heißt  $\mathfrak{I}\mathfrak{m}(X^T) \cap \mathfrak{I}\mathfrak{m}(H^T) = \{\vec{0}\}$ , und  $\text{rg}(G) = k$ .

### Satz 2.1.7.

Es sei  $X$  die  $(n \times k)$ -Designmatrix des durch Gleichung (2.1.1) gegebenen linearen Modells mit  $\text{rg}(X) = p$  ( $p < k$ ) und  $H$  eine  $(m \times k)$ -Matrix. Sind die Gleichungen

$$X\vec{\psi} = X\vec{\beta}, \quad H\vec{\psi} = \vec{0} \quad (2.1.110)$$

für alle  $\vec{\beta} \in \mathbb{R}^k$  eindeutig lösbar, dann ist die durch diese Gleichungen implizit gegebene Funktion  $\vec{\psi} : \vec{\beta} \mapsto \vec{\psi}(\vec{\beta})$  eine schätzbare lineare Funktion mit

$$\vec{\psi}(\vec{\beta}) = (X^T X + H^T H)^{-1} X^T X \vec{\beta} \quad (2.1.111)$$

Die Matrix  $(X^T X + H^T H)^{-1}$  ist eine verallgemeinerte Inverse von  $X^T X$ .

Der BLE-Schätzer

$$\hat{\beta}_* = \hat{\psi} = \vec{\psi}(\hat{\beta}) = (X^T X + H^T H)^{-1} X^T X \hat{\beta}, \quad (2.1.112)$$

wobei  $\hat{\beta}$  eine beliebige Lösung der Normalgleichungen ist, stimmt mit derjenigen (eindeutigen) Lösung  $\tilde{\beta}$  der Normalgleichungen überein, die den Nebenbedingungen (Reparametrisierungsbedingungen)  $H\tilde{\beta} = \vec{0}$  genügt.

**Beweis:**

Die Gleichungen  $X\vec{\psi} = X\vec{\beta}$ ,  $H\vec{\psi} = \vec{0}$  ergeben mit der  $((n+m) \times k)$ -Matrix  $G = \begin{bmatrix} X \\ H \end{bmatrix}$  das Gleichungssystem

$$G\vec{\psi} = \begin{bmatrix} X \\ H \end{bmatrix} \vec{\psi} = \begin{bmatrix} X\vec{\beta} \\ \vec{0} \end{bmatrix} \quad (2.1.113)$$

Ist dieses Gleichungssystem eindeutig lösbar, dann gilt  $rg(G) = rg(G^T G) = k$ , und durch Multiplikation der Gleichung (2.1.113) von links mit  $G^T$  ergibt sich

$$G^T G \vec{\psi} = \begin{bmatrix} X^T & H^T \end{bmatrix} \begin{bmatrix} X \\ H \end{bmatrix} \vec{\psi} = \begin{bmatrix} X^T & H^T \end{bmatrix} \begin{bmatrix} X\vec{\beta} \\ \vec{0} \end{bmatrix}, \quad (2.1.114)$$

hieraus folgt

$$(X^T X + H^T H) \vec{\psi} = X^T X \vec{\beta}. \quad (2.1.115)$$

Da  $rg(G^T G) = rg(G) = k$  gilt, ist  $G^T G = X^T X + H^T H$  regulär, und wir erhalten als Lösung des durch Gleichung (2.1.113) gegebenen Gleichungssystems mit Nebenbedingungen

$$\vec{\psi}(\vec{\beta}) = (X^T X + H^T H)^{-1} X^T X \vec{\beta}.$$

Multiplizieren wir diese Gleichung mit  $X^T X$  von links und berücksichtigen, dass  $X\vec{\psi}(\vec{\beta}) = X\vec{\beta}$  gilt, erhalten wir

$$X^T X \vec{\beta} = X^T X \vec{\psi}(\vec{\beta}) = X^T X (X^T X + H^T H)^{-1} X^T X \vec{\beta} \quad \forall \vec{\beta} \in \mathbb{R}^k.$$

Somit gilt

$$X^T X = X^T X (X^T X + H^T H)^{-1} X^T X, \quad (2.1.116)$$

das heißt, die Matrix  $(X^T X + H^T H)^{-1}$  ist eine verallgemeinerte Inverse von  $X^T X$ .

Nach den Kriterien für Schätzbarkeit ist  $\vec{\psi}$  schätzbar und nach Satz 2.1.6 ist

$$\hat{\psi} = (X^T X + H^T H)^{-1} X^T X \hat{\beta}, \quad (2.1.117)$$

wobei  $\hat{\beta}$  eine beliebige Lösung der Normalgleichungen ist, der BLE-Schätzer für  $\vec{\psi}$ .

Da  $\hat{\psi} = \vec{\psi}(\hat{\beta})$  gilt, ist  $\hat{\psi}$  die eindeutig bestimmte Lösung der Gleichungen

$$\hat{y} = X\hat{\beta} = X\hat{\psi}, \quad H\hat{\psi} = \vec{0} \quad (2.1.118)$$

und somit auch die eindeutig bestimmte Lösung der Normalgleichungen, die die Nebenbedingungen  $H\hat{\beta} = \vec{0}$  erfüllt.

**Beispiel 2.1.4. Einfache Varianzanalyse**

In dem durch Gleichung (2.1.55) gegebenen Modell der einfachen Varianzanalyse ist der Rang der Designmatrix  $X$   $k$  und die Anzahl der Spalten  $k + 1$ . Nach Satz 2.1.7 kann die Eindeutigkeit der Parametrisierung erreicht werden durch Einführung einer linearen Nebenbedingung der Form

$$\vec{h}^T \vec{\beta} = 0, \quad \vec{h} \in \mathbb{R}^{k+1} \quad (2.1.119)$$

mit einem Vektor  $\vec{h}$ , der nicht im Zeilenraum von  $X$  liegt. Nach Gleichung (2.1.98) bilden die Vektoren  $(\sum_{i=1}^k l_i, l_1, l_2, \dots, l_k)^T$ ,  $\vec{l} = (l_1, \dots, l_k)^T \in \mathbb{R}^k$ , den Zeilenraum von  $X$ .

Die Nebenbedingung kann abgeleitet werden aus der Bedeutung, die dem künstlich in das Problem eingebrachten Parameter  $\mu$  durch die Bezeichnung als Gesamtmittel zugewiesen wird, sowie dem Zusammenhang mit dem durch Gleichung (2.1.49) gegebenen ursprünglichen Modell. Naheliegend sind die beiden folgenden Festlegungen von  $\mu$

1.  $\mu = \frac{1}{k} \sum_{i=1}^k \mu_i = \mu + \frac{1}{k} \sum_{i=1}^k \alpha_i$
2.  $\mu = \frac{1}{n} \sum_{i=1}^k n_i \mu_i = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} E[Y_{ij}] = E[\bar{Y}_{..}] = \mu + \frac{1}{n} \sum_{i=1}^k n_i \alpha_i$

Im ersten Fall ist  $\mu$  der Mittelwert der Erwartungswerte der Stufen des zu untersuchenden Faktors. Dieses führt auf die Nebenbedingung

$$\sum_{i=1}^k \alpha_i = 0, \quad \vec{h} = (0, 1, \dots, 1)^T \quad (2.1.120)$$

Im zweiten Fall ist  $\mu$  der Mittelwert der Erwartungswerte aller Beobachtungen. Dieses führt auf die Nebenbedingung

$$\sum_{i=1}^k n_i \alpha_i = 0, \quad \vec{h} = (0, n_1, \dots, n_k)^T \quad (2.1.121)$$

In beiden Fällen liegt der Koeffizientenvektor  $\vec{h}$  nicht im Zeilenraum von  $X$ . Der BLE-Schätzer des Parametervektors, der eine der Nebenbedingungen erfüllt, ist der eindeutig bestimmte Lösungsvektor  $(\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_k)$  der Normalgleichungen, der die Nebenbedingung erfüllt. Da die erste Gleichung der Normalgleichungen (2.1.57) die Summe der übrigen  $k$  Gleichungen ist, verbleiben die Normalgleichungen:

$$n_i \hat{\mu} + n_i \hat{\alpha}_i = Y_{i.}, \quad i = 1, \dots, k \quad (2.1.122)$$

die äquivalent sind zu

$$\hat{\alpha}_i = \bar{Y}_{i.} - \hat{\mu}, \quad i = 1, \dots, k \quad (2.1.123)$$

Im ersten Fall ergibt die Nebenbedingung (2.1.120)

$$0 = \sum_{i=1}^k \hat{\alpha}_i = \sum_{i=1}^k \bar{Y}_{i.} - k \hat{\mu} \quad (2.1.124)$$

*hiermit erhält man die BLE-Schätzer der Komponenten des durch diese Nebenbedingung bestimmten projizierten Parametervektors der einfachen Varianzanalyse*

$$\hat{\mu} = \frac{1}{k} \sum_{i=1}^k \bar{Y}_{i.}, \quad \hat{\alpha}_i = \bar{Y}_{i.} - \frac{1}{k} \sum_{i=1}^k \bar{Y}_{i.} \quad (2.1.125)$$

*Im zweiten Fall ergibt die Nebenbedingung (2.1.121)*

$$0 = \sum_{i=1}^k n_i \hat{\alpha}_i = \sum_{i=1}^k n_i \bar{Y}_{i.} - n \hat{\mu} \quad (2.1.126)$$

*hiermit erhält man die BLE-Schätzer der Komponenten des durch diese Nebenbedingung bestimmten projizierten Parametervektors der einfachen Varianzanalyse*

$$\hat{\mu} = \bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^k n_i \bar{Y}_{i.}, \quad \hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..} \quad (2.1.127)$$


---



## 2.2 Streuungszerlegung und kanonische Darstellung

### Bemerkung 1. Streuungszerlegung für das lineare Modell

Es sei  $\hat{Y} = X\hat{\beta} = \hat{\mu}_{\vec{Y}}$  der BLE-Schätzer des Erwartungswertes des  $n$ -dimensionalen Beobachtungsvektors  $\vec{Y}$  und  $\hat{\beta}$  eine Lösung der Normalgleichungen des linearen Modells

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\beta} \in \mathbb{R}^k. \quad (2.2.1)$$

Da der Beobachtungsvektor  $\vec{Y}$  in zwei orthogonale Komponenten zerfällt,

$$\vec{Y} = \hat{Y} + (\vec{Y} - \hat{Y}), \quad \hat{Y} \perp \vec{Y} - \hat{Y}, \quad (2.2.2)$$

ergibt sich mit dem Satz von Pythagoras für das lineare Modell die sogenannte Streuungszerlegung

$$\|\vec{Y}\|^2 = \|\hat{Y}\|^2 + \|\vec{Y} - \hat{Y}\|^2. \quad (2.2.3)$$

Die totale Quadratsumme

$$SQ_T := \|\vec{Y}\|^2 = \vec{Y}^T \vec{Y} = \sum_{i=1}^n Y_i^2 \quad (2.2.4)$$

ist ein Maß für die Gesamtstreuung des Beobachtungsvektors  $\vec{Y}$  bzw. der  $Y$ -Werte (um 0).

Die Residuen- oder Fehlerquadratsumme

$$SQ_E := \|\vec{Y} - \hat{Y}\|^2 = (\vec{Y} - \hat{Y})^T (\vec{Y} - \hat{Y}) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.2.5)$$

ist ein Maß für die Streuung des Residuenvektors, der Schätzung des Fehlervektors,

$$\hat{\epsilon} := \vec{Y} - \hat{Y} = \vec{Y} - \hat{\mu}_{\vec{Y}}. \quad (2.2.6)$$

Die Modellquadratsumme bzw. Regressionsquadratsumme

$$SQ_R := \|\hat{Y}\|^2 = \hat{Y}^T \hat{Y} = \sum_{i=1}^n \hat{Y}_i^2 \quad (2.2.7)$$

ist ein Maß für die Streuung der Schätzung des Erwartungswertvektors  $\hat{Y} = X\hat{\beta} = \hat{\mu}_{\vec{Y}}$  (um 0). Die Komponenten  $\hat{Y}_i$  von  $\hat{Y}$  bezeichnet man auch als die angepassten bzw. gefitteten  $Y$ -Werte.

Mit diesen Bezeichnungen lautet die Streuungszerlegung (2.2.3)

$$SQ_T = SQ_R + SQ_E. \quad (2.2.8)$$

Da  $\hat{Y} \perp (\vec{Y} - \hat{Y})$  und  $\hat{Y} = X\hat{\beta}$  ergibt sich für  $SQ_R$  die Darstellung

$$SQ_R = \hat{Y}^T \hat{Y} = \hat{Y}^T (\vec{Y} + (\hat{Y} - \vec{Y})) = \hat{Y}^T \vec{Y} = \hat{\beta}^T X^T \vec{Y}. \quad (2.2.9)$$

Man bezeichnet  $\hat{Y}$  als den Anteil von  $\vec{Y}$ , der von den kontrollierten Variablen “erklärt” wird, und  $\hat{\epsilon} = \vec{Y} - \hat{Y}$  als den unerklärten Rest.  $SQ_E = \hat{\epsilon}^T \hat{\epsilon} = SQ_T - SQ_R$  ist die Streuung (Quadratsumme) des unerklärten Restes, die unerklärte Reststreuung, und  $SQ_R$  der Anteil der Gesamtstreuung (totalen Quadratsumme), der durch das lineare Modell erklärt wird.  $SQ_R$  ist die auf das lineare Modell zurückzuführende Reduktion der totalen Quadratsumme.

**Tabelle der Streuungszerlegung (Gesamtstreuung):**

Ursache	Freiheitsgrade FG	Summe der Quadrate SQ	Mittl. Summe der Quadrate MQ	F-Wert
Modell		$SQ_R = \ \hat{Y}\ ^2 = \hat{\beta}^T X^T \vec{Y}$		
Fehler		$SQ_E = \ \vec{Y} - \hat{Y}\ ^2$		
Total		$SQ_T = \vec{Y}^T \vec{Y}$		

**Bemerkung 2. Streuungszerlegung bei gleichem Erwartungswert**

Betrachten wir den Spezialfall von Beobachtungen mit gleichem Erwartungswert:

$$\vec{Y} = \vec{1}_n \mu + \vec{\epsilon}, \quad E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \mu \in \mathbb{R}. \quad (2.2.10)$$

Die BLE-Schätzer für  $\mu$  und  $\vec{\mu}_{\vec{Y}}$  ergeben sich zu

$$\hat{\mu} = \left( \vec{1}_n^T \vec{1}_n \right)^{-1} \vec{1}_n^T \vec{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}, \quad \hat{Y} = \hat{\mu}_{\vec{Y}} = \vec{1}_n \bar{Y} = P_{\mathfrak{Jm}(\vec{1}_n)} \vec{Y} \quad (2.2.11)$$

Für dieses Modell ergibt die Streuungszerlegung (2.2.3)

$$SQ_T = \|\vec{Y}\|^2 = \|\hat{Y}\|^2 + \|\vec{Y} - \hat{Y}\|^2 = n\bar{Y}^2 + \|\vec{Y} - \vec{1}_n \bar{Y}\|^2 \quad (2.2.12)$$

Die Modellquadratsumme dieses speziellen linearen Modells bezeichnen wir als Mittelwertkorrektur

$$SQ_M = n\bar{Y}^2 = \frac{1}{n} \left( \sum_{i=1}^n Y_i \right)^2 \quad (2.2.13)$$

und die Fehlerquadratsumme als (um den Mittelwert) bereinigte bzw. korrigierte totale Quadratsumme

$$SQ_{Tc} = \|\vec{Y} - \vec{1}_n \bar{Y}\|^2 = (\vec{Y} - \vec{1}_n \bar{Y})^T (\vec{Y} - \vec{1}_n \bar{Y}) = \sum_{i=1}^n (Y_i - \bar{Y})^2 = SQ_T - SQ_M. \quad (2.2.14)$$

Mit diesen Bezeichnungen lautet die Streuungszerlegung (2.2.12)

$$SQ_T = SQ_M + SQ_{Tc}. \quad (2.2.15)$$

$SQ_M$  ist die auf den Mittelwert zurückzuführende Reduktion der Gesamtstreuung  $SQ_T$ .

**Bemerkung 3. Streuungszerlegung für das lineare Modell mit Achsenabschnitt**

Wir betrachten im Folgenden das lineare Modell mit Achsenabschnitt für den  $n$ -dimensionalen Beobachtungsvektor  $\vec{Y} = (Y_1, \dots, Y_n)^T$

$$\begin{aligned} Y_i &= \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k + \epsilon_i, \quad i = 1, \dots, n \\ \vec{Y} &= X\vec{\beta} + \vec{\epsilon}, \quad E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T \in \mathbb{R}^{k+1} \\ X &= \begin{bmatrix} \vec{1}_n & X_1 \end{bmatrix}, \quad X_1 = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, k}} = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \dots & \vec{x}_k \end{bmatrix} \\ \bar{x} &= (\bar{x}_1, \dots, \bar{x}_k)^T, \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \end{aligned} \quad (2.2.16)$$

Es sei  $\hat{\beta}$  eine Lösung der Normalgleichungen  $X^T X \hat{\beta} = X^T \vec{Y}$ ,  $\hat{Y} = X \hat{\beta} = \hat{\mu}_{\vec{Y}}$  der BLE-Schätzer des Erwartungswertes von  $\vec{Y}$  und  $\hat{\epsilon} = \vec{Y} - \hat{Y}$  der Residuenvektor. Die erste Normalgleichung lautet

$$\vec{1}_n^T X \hat{\beta} = \vec{1}_n^T \vec{Y} \quad (2.2.17)$$

bzw. nach Division durch  $n$

$$\hat{\beta}_0 + \bar{x}_1 \hat{\beta}_1 + \dots + \bar{x}_k \hat{\beta}_k = \bar{Y} \quad (2.2.18)$$

somit erfüllen die empirischen Mittelwerte die lineare Beziehung exakt, wenn der Parametervektor durch eine Lösung der Normalgleichungen ersetzt wird.

Aus der ersten Normalgleichung (2.2.17) ergibt sich, da  $\hat{Y} = X \hat{\beta}$ , die Beziehung  $\vec{1}_n^T \hat{Y} = \vec{1}_n^T \vec{Y}$  oder  $\bar{\hat{Y}} = \bar{Y}$ , was gleichwertig ist mit  $\bar{\hat{\epsilon}} = 0$ . Somit ist die Summe der Residuen, die Summe der Abweichungen der Beobachtungen von den Werten der geschätzten Modellfunktion, Null.

Da  $\bar{\hat{Y}} = \bar{Y}$  und nach Gleichung (2.2.11)  $\vec{1}_n \bar{\hat{Y}} = P_{\mathfrak{Jm}(\vec{1}_n)} \hat{Y}$ , ergibt sich

$$\vec{1}_n \bar{Y} = \vec{1}_n \bar{\hat{Y}} = P_{\mathfrak{Jm}(\vec{1}_n)} \hat{Y} \in \mathfrak{Jm}(\vec{1}_n) \subseteq \mathfrak{Jm}(X) \quad (2.2.19)$$

$$\hat{Y} - \vec{1}_n \bar{Y} = \hat{Y} - P_{\mathfrak{Jm}(\vec{1}_n)} \hat{Y} \in \mathfrak{Jm}(\vec{1}_n)^\perp \cap \mathfrak{Jm}(X) \quad (2.2.20)$$

$$\vec{Y} - \hat{Y} \in \mathfrak{Jm}(X)^\perp. \quad (2.2.21)$$

Somit sind die Vektoren  $\vec{1}_n \bar{Y}$ ,  $\hat{Y} - \vec{1}_n \bar{Y}$  und  $\vec{Y} - \hat{Y}$  paarweise orthogonal, sie bilden eine orthogonale Zerlegung von  $\vec{Y}$ .

Für die Modellquadratsumme ergibt sich hieraus die Zerlegung

$$SQ_R = \|\hat{Y}\|^2 = \|\vec{1}_n \bar{\hat{Y}}\|^2 + \|\hat{Y} - \vec{1}_n \bar{\hat{Y}}\|^2 = n \bar{Y}^2 + \|\hat{Y} - \vec{1}_n \bar{Y}\|^2 \quad (2.2.22)$$

Die Quadratsumme

$$SQ_{Rc} := \|\hat{Y} - \vec{1}_n \bar{Y}\|^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = SQ_R - SQ_M \quad (2.2.23)$$

bezeichnen wir als die (um den Mittelwert) bereinigte bzw. korrigierte Modellquadratsumme dieses Modells.

Somit ergeben sich für das lineare Modell mit Achsenabschnitt (mit  $\vec{1}_n$  im Spaltenraum von  $X$ ) aus den Gleichungen (2.2.8), (2.2.14) und (2.2.23) die folgenden Streuungszerlegungen:

$$SQ_T = SQ_M + SQ_{Rc} + SQ_E \quad (2.2.24)$$

$$SQ_{Tc} = SQ_{Rc} + SQ_E \quad (2.2.25)$$

Das lineare Modell (2.2.10) ergibt sich aus einem durch Gleichung (2.2.16) gegebenen vollständigen linearen Modells mit Achsenabschnitt durch Einschränkung des Parameterraumes.  $SQ_M$  ist die auf den Mittelwert zurückzuführende Reduktion der Gesamtstreuung  $SQ_T$ .  $SQ_{Rc}$  ist die auf das vollständige lineare Modell zurückzuführende Reduktion der bereinigten Gesamtstreuung  $SQ_{Tc}$  und somit die zusätzliche Reduktion der totalen Quadratsumme  $SQ_T$ , die das vollständige lineare Modell gegenüber dem eingeschränkten Modell ergibt.

**Tabelle der Streuungszerlegung (Bereinigte Gesamtstreuung bei Achsenabschnittsmodell):**

Ursache	Freiheitsgrade FG	Summe der Quadrate SQ	Mittl. Summe der Quadrate MQ	F-Wert
Modell		$SQ_{Rc} = \ \hat{Y} - \bar{Y}\vec{1}_n\ ^2$ $= \hat{\beta}^T X^T \vec{Y} - n\bar{Y}^2$		
Fehler		$SQ_E = \ \vec{Y} - \hat{Y}\ ^2$		
Korr. total		$SQ_{Tc} = \ \vec{Y} - \bar{Y}\vec{1}_n\ ^2$ $= \vec{Y}^T \vec{Y} - n\bar{Y}^2$		

**Definition 2.2.1. Bestimmtheitsmaß für das lineare Modell**

1. Für das durch Gleichung (2.2.1) gegebene lineare Modell mit Achsenabschnitt heißt

$$R^2 := \frac{SQ_{Rc}}{SQ_{Tc}} \quad (2.2.26)$$

Bestimmtheitsmaß.

2. Für das durch Gleichung (2.2.1) gegebene lineare Modell ohne Achsenabschnitt heißt

$$R^2 := \frac{SQ_R}{SQ_T} \quad (2.2.27)$$

Bestimmtheitsmaß.

**Satz 2.2.1.**

Für das Bestimmtheitsmaß eines linearen Modells mit Achsenabschnitt gilt:

1.  $SQ_E = SQ_{Tc} - SQ_{Rc} = SQ_{Tc}(1 - R^2)$
2.  $0 \leq R^2 = 1 - \frac{SQ_E}{SQ_{Tc}} \leq 1$
3.  $R^2 = 1 \Leftrightarrow SQ_E = 0$
4.  $R^2 = 0 \Leftrightarrow SQ_E = SQ_{Tc}$

**Beispiel 2.2.1. Einfache lineare Regression**

Für die in Beispiel 2.1.1 behandelte einfache lineare Regression ergibt sich:

$$\begin{aligned}
 SQ_E &= \sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 = \sum_{i=1}^n (Y_i - (\hat{\alpha} + \hat{\beta}(x_i - \bar{x})))^2 \\
 &= SQ_T - SQ_R = \vec{Y}^T \vec{Y} - \hat{\gamma}^T \tilde{X}^T \vec{Y} = \vec{Y}^T \vec{Y} - \left( \bar{Y}, \frac{s_{xY}}{s_{xx}} \right) \begin{pmatrix} n\bar{Y} \\ ns_{xY} \end{pmatrix} \\
 &= \vec{Y}^T \vec{Y} - n\bar{Y}^2 - n \frac{s_{xY}^2}{s_{xx}} = ns_{YY} - n \frac{s_{xY}^2}{s_{xx}} \\
 &= ns_{YY}(1 - r_{xY}^2) \\
 SQ_{Tc} &= ns_{YY}, \quad SQ_{Rc} = ns_{\hat{Y}\hat{Y}} = n \frac{s_{xY}^2}{s_{xx}} = ns_{YY} r_{xY}^2 = \hat{\beta} ns_{xY} \\
 R^2 &= r_{xY}^2
 \end{aligned} \quad (2.2.28)$$

**Beispiel 2.2.2. Einfache Varianzanalyse**

Für die in Beispiel 2.1.2 behandelte einfache Varianzanalyse ergibt sich:

$$\begin{aligned}
 SQ_{Tc} &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - n\bar{Y}_{..}^2 \\
 SQ_E &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 = SQ_{in} \\
 &= SQ_T - SQ_R = \vec{Y}^T \vec{Y} - \hat{\gamma}^T \tilde{X}^T \vec{Y} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^k n_i \bar{Y}_{i.}^2 \\
 SQ_{Rc} &= \sum_{i=1}^k \sum_{j=1}^{n_i} (\hat{Y}_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 = SQ_{zw} \\
 SQ_{Tc} &= SQ_{in} + SQ_{zw}
 \end{aligned} \tag{2.2.29}$$

Die Gesamtstreuung der Daten um den Mittelwert aller Daten,  $SQ_{Tc}$ , wird zerlegt in die Streuung der Daten innerhalb der Gruppen,  $SQ_{in}$ , und die Streuung zwischen den Gruppen,  $SQ_{zw}$ .

**Definition 2.2.2. Kanonische Basis**

Eine Orthonormalbasis  $\vec{q}_1, \dots, \vec{q}_n$  des  $\mathbb{R}^n$  heißt kanonische Basis für das lineare Modell

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\epsilon}} = \sigma^2 I, \quad \vec{\beta} \in \mathbb{R}^k, \tag{2.2.30}$$

dessen Designmatrix  $X$  den Rang  $r$  ( $r \leq k$ ) hat, genau dann, wenn das System  $\vec{q}_1, \dots, \vec{q}_r$  eine Basis des Spaltenraumes von  $X$  bildet. Die Darstellung eines linearen Modells bezüglich einer kanonischen Basis heißt kanonische Darstellung des linearen Modells.

**Satz 2.2.2.**

Es sei  $\vec{q}_1, \dots, \vec{q}_n$  eine kanonische Basis des durch Gleichung (2.2.30) gegebenen linearen Modells. Des weiteren sei

$$Q = \begin{bmatrix} \vec{q}_1 & \dots & \vec{q}_n \end{bmatrix}, \quad Q_1 = \begin{bmatrix} \vec{q}_1 & \dots & \vec{q}_r \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \vec{q}_{r+1} & \dots & \vec{q}_n \end{bmatrix}. \tag{2.2.31}$$

Es sei  $\vec{Z}$  die Koordinatendarstellung von  $\vec{Y}$  bzgl. der kanonischen Basis,

$$\vec{Y} = \sum_{i=1}^n \vec{q}_i Z_i = Q\vec{Z}, \tag{2.2.32}$$

dann folgt für  $\vec{Z}$

$$\vec{Z} = Q^T \vec{Y} = \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} X\vec{\beta} + Q^T \vec{\epsilon} = \begin{bmatrix} Q_1^T X\vec{\beta} \\ \vec{0} \end{bmatrix} + Q^T \vec{\epsilon}. \tag{2.2.33}$$

Hieraus ergibt sich mit  $\vec{\gamma} := Q_1^T X \vec{\beta}$ , der eindeutigen Koordinatendarstellung von  $E[\vec{Y}] = X \vec{\beta}$  bezüglich der Basis  $\vec{q}_1, \dots, \vec{q}_r$  des Spaltenraumes von  $X$

$$X \vec{\beta} = Q Q^T X \vec{\beta} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \vec{\gamma} \\ \vec{0} \end{bmatrix} = Q_1 \vec{\gamma} = \sum_{i=1}^r \vec{q}_i \gamma_i, \quad (2.2.34)$$

und  $\vec{\eta} := Q^T \vec{\epsilon}$  die kanonische Darstellung des durch Gleichung (2.2.30) gegebenen linearen Modells bezüglich der gegebenen kanonischen Basis zu

$$\vec{Z} = \begin{bmatrix} \vec{\gamma} \\ \vec{0} \end{bmatrix} + \vec{\eta}, \quad E[\vec{\eta}] = Q^T E[\vec{\epsilon}] = \vec{0}, \quad \Sigma_{\vec{\eta}} = Q^T \Sigma_{\vec{\epsilon}} Q = \sigma^2 I, \quad \vec{\gamma} \in \mathbb{R}^r. \quad (2.2.35)$$

Da  $\hat{Y}$  die orthogonale Projektion auf den Spaltenraum von  $X$  und somit auf den Spaltenraum von  $Q_1$  und  $\hat{\epsilon} = \vec{Y} - \hat{Y}$  die orthogonale Projektion auf das orthogonale Komplement des Spaltenraumes von  $X$  und somit auf den Spaltenraum von  $Q_2$  ist, ergibt sich mit  $\vec{Z}_1 = (Z_1, \dots, Z_r)^T = Q_1^T \vec{Y}$  und  $\vec{Z}_2 = (Z_{r+1}, \dots, Z_n)^T = Q_2^T \vec{Y}$  (siehe Korollar 1.2 zu Definition 1.2.3)

$$\hat{Y} = Q_1 \vec{Z}_1 = \sum_{i=1}^r \vec{q}_i Z_i, \quad \hat{\epsilon} = \vec{Y} - \hat{Y} = Q_2 \vec{Z}_2 = \sum_{i=r+1}^n \vec{q}_i Z_i \quad (2.2.36)$$

$$\hat{\epsilon}^T \hat{\epsilon} = (\vec{Y} - \hat{Y})^T (\vec{Y} - \hat{Y}) = \vec{Z}_2^T Q_2^T Q_2 \vec{Z}_2 = \vec{Z}_2^T \vec{Z}_2 = \sum_{i=r+1}^n Z_i^2 \quad (2.2.37)$$

$$E[\hat{\epsilon}^T \hat{\epsilon}] = E[(\vec{Y} - \hat{Y})^T (\vec{Y} - \hat{Y})] = \sum_{i=r+1}^n E[Z_i^2] = (n - r) \sigma^2 \quad (2.2.38)$$

### Satz 2.2.3.

Es sei  $\hat{\mu}_{\vec{Y}}$  der BLE-Schätzer des Erwartungswertes des  $n$ -dimensionalen Beobachtungsvektors  $\vec{Y}$  und  $\hat{\beta}$  eine Lösung der Normalgleichungen des durch Gleichung (2.2.30) gegebenen linearen Modells mit  $\text{rg}(X) = r < n$ . Dann ist (s. Gl. (2.1.6), (2.2.8), (2.2.9))

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n - r} \|\vec{Y} - \hat{\mu}_{\vec{Y}}\|^2 = \frac{1}{n - r} \|\vec{Y} - X \hat{\beta}\|^2 \\ &= \frac{1}{n - r} (\vec{Y}^T \vec{Y} - \hat{\beta}^T X^T \vec{Y}) \\ &= \frac{1}{n - r} \vec{Y}^T (I - P_{\mathcal{M}(X)}) \vec{Y} \end{aligned} \quad (2.2.39)$$

eine erwartungstreue Schätzung der Varianz  $\sigma^2$ .

## Kapitel 3

# Das klassische lineare Modell

### 3.1 Die mehrdimensionale Normalverteilung

#### Definition 3.1.1. Eindimensionale Normalverteilung

1. Eine Zufallsgröße  $X$  mit der Dichte

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp \left\{ -\frac{(x - \mu)^2}{2 \cdot \sigma^2} \right\} \quad -\infty < x < \infty \quad (3.1.1)$$

mit  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  heißt normalverteilt mit den Parametern  $\mu$  und  $\sigma^2$ :  $X \sim \mathfrak{N}(\mu, \sigma^2)$ .

2. Eine normalverteilte Zufallsgröße mit den Parametern  $\mu = 0$ ,  $\sigma = 1$  heißt standardisiert normalverteilt mit der Dichte

$$\phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}} \quad -\infty < x < \infty \quad (3.1.2)$$

und der Verteilungsfunktion

$$\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^x e^{-\frac{\xi^2}{2}} d\xi \quad -\infty < x < \infty \quad (3.1.3)$$

#### Satz 3.1.1. Eindimensionale Normalverteilung

Es sei  $U$  eine standardisiert normalverteilte Zufallsgröße und  $X = \mu + \sigma U$  mit  $\mu \in \mathbb{R}$  und  $\sigma \in \mathbb{R}_{>0}$ .

1. Für die standardisiert normalverteilte Zufallsgröße  $U$  gilt:

$$E[U] = E[U^3] = 0, \quad \text{Var}(U) = E[U^2] = 1, \quad E[U^4] = 3, \quad (3.1.4)$$

$$\text{Var}(U^2) = E[U^4] - E[U^2]^2 = 2, \quad \text{Kov}(U, U^2) = E[U^3] - E[U]E[U^2] = 0 \quad (3.1.5)$$

$$E[U^{2n-1}] = 0, \quad E[U^{2n}] = \prod_{i=1}^n (2i-1) \quad \forall n \in \mathbb{N}_{>0} \quad (3.1.6)$$

$$\begin{aligned} m_U(t) &= E[e^{tU}] = \int_{\mathbb{R}} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{u^2}{2} + tu} du = e^{\frac{t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{(u-t)^2}{2}} du \\ &= e^{\frac{t^2}{2}} \quad \forall t \in \mathbb{R} \end{aligned} \quad (3.1.7)$$

2. Für die  $\mathfrak{N}(\mu, \sigma^2)$ -normalverteilte Zufallsgröße  $X$  gilt:

$$E[X] = \mu, \quad \text{Var}(X) = \sigma^2 \quad (3.1.8)$$

$$\begin{aligned} m_X(t) &= E[e^{tX}] = E[e^{t(\mu + \sigma U)}] = e^{t\mu} E[e^{(t\sigma)U}] = e^{t\mu} m_U(t\sigma) \\ &= e^{t\mu + \frac{t^2 \sigma^2}{2}} \quad \forall t \in \mathbb{R} \end{aligned} \quad (3.1.9)$$

**Definition 3.1.2. Mehrdimensionale (multivariate) Normalverteilung**

Es sei  $\vec{U} = (U_1, \dots, U_k)^T$  ein Zufallsvektor, dessen Komponenten  $U_1, \dots, U_k$  unabhängige standardisiert normalverteilte Zufallsgrößen sind,  $A$  eine  $(p \times k)$ -Matrix und  $\vec{a} \in \mathbb{R}^p$ . Dann heißt die Verteilung des  $p$ -dimensionalen Zufallsvektors  $\vec{X} = A\vec{U} + \vec{a}$  mehrdimensionale, multivariate oder  $p$ -dimensionale Normalverteilung.

**Satz 3.1.2. Mehrdimensionale Normalverteilung**

Es seien  $\vec{U}, \vec{X}, A = [\vec{a}_1 \quad \vec{a}_2 \quad \dots \quad \vec{a}_k]$  und  $\vec{a}$  definiert wie in Definition 3.1.2. Dann gilt:

$$\vec{\mu} = \vec{\mu}_{\vec{X}} = E[\vec{X}] = E[A\vec{U} + \vec{a}] = AE[\vec{U}] + \vec{a} = \vec{a} \quad (3.1.10)$$

$$\Sigma = \Sigma_{\vec{X}} = \text{Kov}(\vec{X}) = \text{Kov}(A\vec{U} + \vec{a}) = A\text{Kov}(\vec{U})A^T = AI_kA^T = AA^T \quad (3.1.11)$$

$$\text{rg}(\Sigma) = \text{rg}(A) \quad (3.1.12)$$

$$\begin{aligned} m_{\vec{X}}(\vec{t}) &= E[e^{\vec{t}^T(A\vec{U} + \vec{a})}] = e^{\vec{t}^T \vec{a}} E[e^{\sum_{i=1}^k (\vec{t}^T \vec{a}_i) U_i}] = e^{\vec{t}^T \vec{a}} \prod_{i=1}^k E[e^{(\vec{t}^T \vec{a}_i) U_i}] = e^{\vec{t}^T \vec{a}} \prod_{i=1}^k m_{U_i}(\vec{t}^T \vec{a}_i) \\ &= e^{\vec{t}^T \vec{a}} \prod_{i=1}^k e^{\frac{1}{2} \vec{t}^T \vec{a}_i \vec{a}_i^T \vec{t}} = e^{\vec{t}^T \vec{a} + \frac{1}{2} \vec{t}^T (\sum_{i=1}^k \vec{a}_i \vec{a}_i^T) \vec{t}} = e^{\vec{t}^T \vec{a} + \frac{1}{2} \vec{t}^T AA^T \vec{t}} \\ &= e^{\vec{t}^T \vec{\mu} + \frac{1}{2} \vec{t}^T \Sigma \vec{t}} \quad \forall \vec{t} \in \mathbb{R}^p \end{aligned} \quad (3.1.13)$$

**Bemerkung 1.**

Es seien  $\vec{U}, \vec{X}, A$  und  $\vec{a}$  definiert wie in Definition 3.1.2.

1. Eine multivariate Normalverteilung wird eindeutig durch den Erwartungswert und die Kovarianzmatrix bestimmt. Daher wird die Schreibweise

$$\vec{X} \sim \mathfrak{N}(\vec{\mu}, \Sigma) \quad (3.1.14)$$

verwendet.

2.  $\vec{U}$  ist ein Zufallsvektor, dessen Komponenten unabhängige standardisiert normalverteilte Zufallsgrößen sind, genau dann, wenn  $\vec{U} \sim \mathfrak{N}(\vec{0}, I_k)$ .
3.  $\Sigma$  ist positiv definit, genau dann wenn  $\text{rg}(A) = p$ .



**Korollar 1.**

Es sei  $\vec{\mu} \in \mathbb{R}^p$  und  $\Sigma$  eine positiv semidefinite Matrix der Ordnung  $p$ . Ist  $\text{rg}(\Sigma) = r > 0$ , dann gibt es eine  $(p \times r)$ -Matrix  $A$  vom Rang  $r$  mit  $\Sigma = AA^T$  (siehe Bemerkung 1 zu Satz 1.3.7). Ist  $\vec{U} = (U_1, \dots, U_r)^T$  ein Zufallsvektor, dessen Komponenten  $U_1, \dots, U_r$  unabhängige standardisiert normalverteilte Zufallsgrößen sind, dann gilt:

$$\vec{X} = A\vec{U} + \vec{\mu} \sim \mathfrak{N}(\vec{\mu}, \Sigma) \quad (3.1.15)$$

**Satz 3.1.3. Mehrdimensionale Normalverteilung**

Es sei  $\vec{X}$  ein  $\mathfrak{N}(\vec{\mu}, \Sigma)$ -verteilter  $p$ -dimensionaler Zufallsvektor. Ist  $B$  eine  $(q \times p)$ -Matrix und  $\vec{b} \in \mathbb{R}^q$ , so gilt

$$\vec{Y} = B\vec{X} + \vec{b} \sim \mathfrak{N}(B\vec{\mu} + \vec{b}, B\Sigma B^T). \quad (3.1.16)$$

Insbesondere sind alle (auch die mehrdimensionalen) Randverteilungen einer Normalverteilung wieder Normalverteilungen.

**Beweis:**

Die Aussage ergibt sich direkt aus der Definition der multivariaten Normalverteilung (Def. 3.1.2).

**Bemerkung 1.**

Die eindimensionalen Randverteilungen einer multivariaten Normalverteilung sind wieder Normalverteilungen. Die Umkehrung dieser Aussage ist nicht richtig. Aus der Tatsache, dass die Komponenten eines Zufallsvektors normalverteilt sind, kann im allgemeinen nicht auf eine gemeinsame Normalverteilung geschlossen werden.

**Bemerkung 2.**

Sind alle Linearformen  $\vec{c}^T \vec{X}$ ,  $\vec{c} \in \mathbb{R}^p$  des  $p$ -dimensionalen Zufallsvektors  $\vec{X}$  normalverteilt, dann ist auch  $\vec{X}$  normalverteilt.

**Satz 3.1.4. Mehrdimensionale Normalverteilung**

1. Es sei  $\vec{X} = (X_1, \dots, X_p)^T$  ein  $\mathfrak{N}(\vec{\mu}, \Sigma)$ -verteilter  $p$ -dimensionaler Zufallsvektor. Sind die Komponenten  $X_1, \dots, X_p$  paarweise unkorreliert, das heißt  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ , dann sind sie auch unabhängig.
2. Sind die Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_l$  gemeinsam normalverteilt, und sind sie paarweise unkorreliert, das heißt  $\Sigma_{\vec{X}_i, \vec{X}_j} = \text{Cov}(\vec{X}_i, \vec{X}_j) = O$  für alle  $i \neq j$ , dann sind sie auch unabhängig.

**Beweis:**

Die Aussage des ersten Teils des Satzes ist ein Spezialfall der Aussage des zweiten Teils, daher reicht es diesen zu beweisen. Es sei

$$\vec{X} = \begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \\ \vdots \\ \vec{X}_l \end{bmatrix}, \quad \vec{\mu} = \begin{bmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \\ \vdots \\ \vec{\mu}_l \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & O & \dots & O \\ O & \Sigma_2 & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & \Sigma_l \end{bmatrix}, \quad \vec{t} = \begin{bmatrix} \vec{t}_1 \\ \vec{t}_2 \\ \vdots \\ \vec{t}_l \end{bmatrix}. \quad (3.1.17)$$

Die  $\vec{X}_i$  ( $i = 1, \dots, l$ ) sind  $d_i$ -dimensionale Zufallsvektoren mit  $E[\vec{X}_i] = \vec{\mu}_i$ ,  $Kov(\vec{X}_i) = \Sigma_i$  und  $\vec{t}_i \in \mathbb{R}^{d_i}$ . Da  $\vec{X} \mathfrak{N}(\vec{\mu}, \Sigma)$ -verteilt ist, ergibt sich die momentenerzeugende Funktion von  $\vec{X}$ , die gemeinsame momentenerzeugende Funktion der Zufallsvektoren  $\vec{X}_1, \dots, \vec{X}_l$ , mit Gleichung (3.1.13) zu

$$\begin{aligned} m_{\vec{X}}(\vec{t}) &= e^{\vec{t}^T \vec{\mu} + \frac{1}{2} \vec{t}^T \Sigma \vec{t}} = e^{\sum_{i=1}^l \vec{t}_i^T \vec{\mu}_i + \frac{1}{2} \sum_{i=1}^l \vec{t}_i^T \Sigma_i \vec{t}_i} = \prod_{i=1}^l e^{\vec{t}_i^T \vec{\mu}_i + \frac{1}{2} \vec{t}_i^T \Sigma_i \vec{t}_i} \\ &= \prod_{i=1}^l m_{\vec{X}_i}(\vec{t}_i) \end{aligned} \quad (3.1.18)$$

Hieraus folgt mit Satz 1.3.5 die behauptete Unabhängigkeit der Zufallsvektoren.

**Satz 3.1.5. Unabhängigkeit von Linearformen normalverteilter Zufallsgrößen**

Der  $n$ -dimensionale Zufallsvektor  $\vec{Y}$  sei  $\mathfrak{N}(\vec{\mu}, \sigma^2 I_n)$ -verteilt,  $\sigma^2 > 0$ . Die Zufallsvariablen  $U_i = \vec{c}_i^T \vec{Y}$ ,  $\vec{c}_i \in \mathbb{R}^n$ ,  $i = 1, \dots, k$ , sind stochastisch unabhängig genau dann, wenn die Vektoren  $\vec{c}_1, \dots, \vec{c}_k$  orthogonal sind.

**Beweis:**

Die Zufallsgrößen  $U_1, \dots, U_k$  besitzen eine gemeinsame Normalverteilung und sind somit unabhängig genau dann, wenn sie paarweise unkorreliert sind. Daher ist zu zeigen, dass die Zufallsgrößen genau dann paarweise unkorreliert sind, wenn die Vektoren  $\vec{c}_i \in \mathbb{R}^n$ ,  $i = 1, \dots, k$  orthogonal sind. Es sei  $i \neq j$ , dann gilt

$$Kov(U_i, U_j) = Kov(\vec{c}_i^T \vec{Y}, \vec{c}_j^T \vec{Y}) = \vec{c}_i^T \Sigma_Y \vec{c}_j = \sigma^2 \vec{c}_i^T \vec{c}_j \quad (3.1.19)$$

und, da  $\sigma^2 > 0$  vorausgesetzt war, folgt hieraus die Behauptung.

**Satz 3.1.6. Mehrdimensionale Normalverteilung**

Es sei  $\vec{\mu} \in \mathbb{R}^p$  und  $\Sigma$  eine positiv semidefinite Matrix der Ordnung  $p$ . Die  $\mathfrak{N}(\vec{\mu}, \Sigma)$ -Verteilung ist stetig, das heißt es existiert eine Dichte, genau dann, wenn  $\Sigma$  positiv definit ist, das heißt, wenn die Verteilung nichtsingulär ist. In diesem Fall stellt

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu})} \quad (3.1.20)$$

eine Dichte der  $\mathfrak{N}(\vec{\mu}, \Sigma)$ -Verteilung dar.

**Beweis:**

Es sei  $\Sigma$  positiv definit. Dann gibt es eine reguläre Matrix  $A$  mit  $\Sigma = AA^T$  (siehe Bemerkung 1 zu Satz 1.3.7). Ist  $\vec{U} \mathfrak{N}(\vec{0}, I_p)$ -verteilt, so ist  $\vec{X} = A\vec{U} + \vec{\mu} \mathfrak{N}(\vec{\mu}, \Sigma)$ -verteilt. Da die Komponenten von  $\vec{U}$  unabhängige standardisiert normalverteilte Zufallsgrößen sind, ist  $\vec{U}$  nach Satz 1.3.2 stetig verteilt mit der Dichte

$$\begin{aligned} f_{\vec{U}}(\vec{u}) &= f_{U_1, \dots, U_p}(u_1, \dots, u_p) = \prod_{i=1}^p \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} u_i^2} \\ &= \frac{1}{(2 \cdot \pi)^{\frac{p}{2}}} \cdot e^{-\frac{1}{2} \vec{u}^T \vec{u}} \quad \forall \vec{u} \in \mathbb{R}^p \end{aligned} \quad (3.1.21)$$

Da  $A$ , die Funktionalmatrix der Abbildung  $g(\vec{u}) = A\vec{u} + \vec{\mu}$  auf dem  $\mathbb{R}^p$ , regulär ist, folgt aus dem Transformationssatz für Dichten (Satz 1.3.3), dass  $\vec{X}$  mit der Dichte

$$\begin{aligned} f_{\vec{X}}(\vec{x}) &= f_{\vec{U}}(A^{-1}(\vec{x} - \vec{\mu})) \cdot |\det(A^{-1})| = \frac{1}{(2 \cdot \pi)^{\frac{p}{2}} |\det(A)|} \cdot e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T (A^{-1})^T A^{-1} (\vec{x} - \vec{\mu})} \\ &= \frac{1}{(2 \cdot \pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})} \quad \forall \vec{x} \in \mathbb{R}^p \end{aligned} \quad (3.1.22)$$

stetig verteilt ist.

Besitzt eine Verteilung eine Dichte, so kann sie nicht auf einen echten affinen Teilraum des  $\mathbb{R}^p$  konzentriert sein, da jeder solcher eine Nullmenge bezüglich des  $p$ -dimensionalen Lebesgue-Maßes darstellt. Hieraus folgt mit Satz 1.3.7 (2.), dass die Kovarianzmatrix der Verteilung regulär ist.

### 3.2 $\chi^2$ -, F- und t-Verteilung

#### Definition 3.2.1. $\chi^2$ -Verteilung

1. Es sei  $\vec{Y} = (Y_1, Y_2, \dots, Y_n)^T$  ein  $\mathfrak{N}(\vec{0}, I)$ -verteilter Zufallsvektor. Dann heißt die Verteilung von  $Z = \sum_{i=1}^n Y_i^2$  (zentrale)  $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden

Schreibweise:

$$Z \sim \chi_n^2 \quad , \quad \chi_{n;\alpha}^2 := \alpha\text{-Quantil der } \chi_n^2\text{-Verteilung}$$

2. Es sei  $\vec{Y} = (Y_1, Y_2, \dots, Y_n)^T$  ein  $\mathfrak{N}(\vec{\mu}, I)$ -verteilter Zufallsvektor. Dann heißt die Verteilung von  $Z = \sum_{i=1}^n Y_i^2$  nichtzentrale  $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden und Nichtzentralitätsparameter  $\delta = \vec{\mu}^T \vec{\mu}$ .

Schreibweise:

$$Z \sim \chi_{n,\delta}^2 \quad , \quad \chi_{n,\delta;\alpha}^2 := \alpha\text{-Quantil der } \chi_{n,\delta}^2\text{-Verteilung}$$

#### Bemerkung 1.

Die für  $\alpha > 0$  durch

$$\Gamma(\alpha) := \int_0^\infty x^{(\alpha-1)} e^{-x} dx \quad (3.2.1)$$

definierte Funktion  $\Gamma$  heißt Gammafunktion.

Mittels partieller Integration erhält man

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \quad (\alpha > 1) \quad , \quad (3.2.2)$$

$$\Gamma(k) = (k - 1)! \quad (k \in \mathbb{N}_{>0}) \quad . \quad (3.2.3)$$

Außerdem gilt

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad (3.2.4)$$

#### Satz 3.2.1. $\chi^2$ -Verteilung

1. Es sei  $Z$  eine  $\chi_n^2$ -verteilte Zufallsgröße. Das heißt, es kann angenommen werden, dass unabhängige  $\mathfrak{N}(0, 1)$ -verteilte Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$  mit  $Z = \sum_{i=1}^n Y_i^2$  existieren. Dann gilt:

$$E[Z] = \sum_{i=1}^n E[Y_i^2] = n \quad (3.2.5)$$

$$\begin{aligned} \text{Var}(Z) &= \sum_{i=1}^n \text{Var}(Y_i^2) = \sum_{i=1}^n \left( \underbrace{E[Y_i^4]}_{=3} - \underbrace{E[Y_i^2]^2}_{=1} \right) \\ &= 2n \end{aligned} \quad (3.2.6)$$

$$m_Z(t) = E[e^{Zt}] = \frac{1}{(1 - 2t)^{\frac{n}{2}}} \quad , \quad t < \frac{1}{2} \quad (3.2.7)$$

$$f_Z(z) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} z^{\frac{(n-1)}{2}} e^{-\frac{z}{2}} & \text{für } z \geq 0 \\ 0 & \text{für } z < 0 \end{cases} . \quad (3.2.8)$$

2. Additionstheorem für die  $\chi_n^2$ -Verteilung:

Es seien  $Z_1$  und  $Z_2$  unabhängige  $\chi_{n_1}^2$ - und  $\chi_{n_2}^2$ -verteilte Zufallsvariablen. Das heißt, es kann angenommen werden, daß unabhängige  $\mathfrak{N}(0, 1)$ -verteilte Zufallsvariablen  $Y_1, \dots, Y_{n_1}, Y_{n_1+1}, \dots, Y_{n_1+n_2}$  mit  $Z_1 := \sum_{i=1}^{n_1} Y_i^2$  und  $Z_2 := \sum_{i=n_1+1}^{n_1+n_2} Y_i^2$  existieren. Dann ist  $Z_1 + Z_2 = \sum_{i=1}^{n_1+n_2} Y_i^2$   $\chi_{n_1+n_2}^2$ -verteilt.

3. Es sei  $Z_n$   $\chi_n^2$ -verteilt. Dann ist

$$U_n = \frac{Z_n - n}{\sqrt{2n}} \quad (3.2.9)$$

asymptotisch normiert normalverteilt. Für die Verteilungsfunktion der  $U_n$  gilt:

$$\lim_{n \rightarrow \infty} F_{U_n}(u) = \Phi(u) \quad (3.2.10)$$

**Beweis:**

Es sei  $U \sim \mathfrak{N}(\mu, 1)$ . Für die momentenerzeugende Funktion von  $U^2$  ergibt sich:

$$\begin{aligned} m_{U^2}(t) &= E[e^{tU^2}] = \int_{-\infty}^{\infty} e^{tu^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-\mu)^2} du \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1-2t}{2}(u^2 - 2\frac{\mu}{1-2t}u + (\frac{\mu}{1-2t})^2) + \frac{1}{2}(\frac{\mu^2}{1-2t} - \mu^2)} du \\ &= \frac{1}{(1-2t)^{\frac{1}{2}}} e^{\frac{t\mu^2}{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}(1-2t)^{-\frac{1}{2}}} e^{-\frac{1}{2}\frac{(u-\tilde{\mu})^2}{(1-2t)^{-1}}} du, \quad \tilde{\mu} = \frac{\mu}{1-2t} \\ &= \frac{1}{(1-2t)^{\frac{1}{2}}} e^{\frac{t\mu^2}{1-2t}}, \quad t < \frac{1}{2}. \end{aligned} \quad (3.2.11)$$

Als nächstes bestimmen wir die momentenerzeugende Funktion der  $\chi_n^2$ -verteilten Zufallsgröße  $Z = \sum_{i=1}^n Y_i^2$ ,  $Y_i \sim \mathfrak{N}(0, 1)$ . Da aus Gleichung (3.2.11)

$$m_{Y_i^2} = E[e^{tY_i^2}] = \frac{1}{(1-2t)^{\frac{1}{2}}} \quad (3.2.12)$$

folgt, ergibt sich mit Satz 1.3.2(1) und Satz 1.3.4

$$m_Z(t) = E[e^{tZ}] = \prod_{i=1}^n E[e^{tY_i^2}] = \frac{1}{(1-2t)^{\frac{n}{2}}} \quad (3.2.13)$$

Dass die in Gleichung (3.2.8) gegebene Funktion  $f_Z$  die Dichte der  $\chi_n^2$ -Verteilung ist, beweisen wir, indem wir zeigen, dass die durch diese Dichte gegebene Verteilung dieselbe momentenerzeugende

Funktion wie  $Z$  hat (s. Satz 1.3.5). Die momentenerzeugende Funktion  $m_{f_Z}$  der durch die Dichte  $f_Z$  gegebenen Verteilung ergibt sich zu

$$\begin{aligned}
 m_{f_Z}(t) &= \int_0^\infty e^{tz} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} z^{(\frac{n}{2}-1)} e^{-\frac{z}{2}} dz \\
 &= \int_0^\infty \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} z^{(\frac{n}{2}-1)} e^{-(\frac{1}{2}-t)z} dz \\
 &= \int_0^\infty \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \left( \frac{x}{\frac{1}{2}-t} \right)^{(\frac{n}{2}-1)} e^{-x} \frac{1}{\frac{1}{2}-t} dx, \quad x = \left( \frac{1}{2}-t \right) z, \quad t < \frac{1}{2} \\
 &= \frac{1}{(1-2t)^{\frac{n}{2}}} \cdot \frac{1}{\Gamma(\frac{n}{2})} \int_0^\infty x^{(\frac{n}{2}-1)} e^{-x} dx = \frac{1}{(1-2t)^{\frac{n}{2}}}
 \end{aligned} \tag{3.2.14}$$

### Satz 3.2.2. Nichtzentrale $\chi^2$ -Verteilung

1. Es sei  $Z$  eine  $\chi_{n,\delta}^2$ -verteilte Zufallsgröße. Das heißt, es kann angenommen werden, dass es einen  $\mathfrak{N}(\vec{\mu}, I)$ -verteilten Zufallsvektor  $\vec{Y} = (Y_1, Y_2, \dots, Y_n)^T$  mit  $Z = \sum_{i=1}^n Y_i^2$  und  $\delta = \vec{\mu}^T \vec{\mu}$  gibt. Dann gilt:

$$m_Z(t) = E[e^{Zt}] = \frac{1}{(1-2t)^{\frac{n}{2}}} e^{\frac{t\delta}{1-2t}}, \quad t < \frac{1}{2} \tag{3.2.15}$$

$$f_Z(z) = \begin{cases} \sum_{j=0}^\infty \frac{(\frac{\delta}{2})^j}{j!} e^{-\frac{\delta}{2}} \frac{1}{2^{\frac{n+2j}{2}} \Gamma(\frac{n+2j}{2})} z^{(\frac{n+2j}{2}-1)} e^{-\frac{z}{2}} & \text{für } z \geq 0 \\ 0 & \text{für } z < 0 \end{cases}. \tag{3.2.16}$$

2. Es sei  $Z$  eine  $\chi_{n,\delta}^2$ -verteilte Zufallsgröße. Das heißt, es kann angenommen werden, dass es einen  $\mathfrak{N}(\vec{0}, I)$ -verteilten Zufallsvektor  $\vec{Y} = (Y_1, Y_2, \dots, Y_n)^T$  mit  $Z = (Y_1 + \sqrt{\delta})^2 + \sum_{i=2}^n Y_i^2$  gibt. Dann gilt:

$$E[Z] = E[(Y_1 + \sqrt{\delta})^2] + \sum_{i=2}^n E[Y_i^2] = (1 + \delta) + (n-1) = n + \delta \tag{3.2.17}$$

$$\begin{aligned}
 \text{Var}(Z) &= \text{Var}((Y_1 + \sqrt{\delta})^2) + \sum_{i=2}^n \text{Var}(Y_i^2) \\
 &= \text{Var}(Y_1^2 + 2Y_1\sqrt{\delta} + \delta) + \sum_{i=2}^n \text{Var}(Y_i^2) \\
 &= \text{Var}(Y_1^2) + 4\delta \text{Var}(Y_1) + 4\sqrt{\delta} \text{Cov}(Y_1^2, Y_1) + \sum_{i=2}^n \text{Var}(Y_i^2) \\
 &= 4\delta + \sum_{i=1}^n \left( \underbrace{E[Y_i^4]}_{=3} - \underbrace{E[Y_i^2]^2}_{=1} \right) \\
 &= 2n + 4\delta
 \end{aligned} \tag{3.2.18}$$

### 3. Additionstheorem für die nichtzentrale $\chi^2$ -Verteilung:

Es seien  $Z_1$  und  $Z_2$  unabhängige  $\chi^2_{n_1, \delta_1}$ - und  $\chi^2_{n_2, \delta_2}$ -verteilte Zufallsvariablen. Das heißt, es kann angenommen werden, dass unabhängige normalverteilte Zufallsvariablen  $Y_1, \dots, Y_{n_1}, Y_{n_1+1}, \dots, Y_{n_1+n_2}$  mit  $\text{Var}(Y_i) = 1, i = 1, \dots, n_1 + n_2, \delta_1 = E[Y_1]^2 + \dots + E[Y_{n_1}]^2, \delta_2 = E[Y_{n_1+1}]^2 + \dots + E[Y_{n_1+n_2}]^2, Z_1 := \sum_{i=1}^{n_1} Y_i^2$  und  $Z_2 := \sum_{i=n_1+1}^{n_1+n_2} Y_i^2$  existieren. Dann ist  $Z_1 + Z_2 = \sum_{i=1}^{n_1+n_2} Y_i^2$   $\chi^2_{n_1+n_2, \delta_1+\delta_2}$ -verteilt.

#### Beweis:

Wir bestimmen die momentenerzeugende Funktion der  $\chi^2_{n, \delta}$ -verteilten Zufallsgröße  $Z = \sum_{i=1}^n Y_i^2, Y_i \sim \mathfrak{N}(\mu_i, 1)$ . Da aus Gleichung (3.2.11)

$$m_{Y_i^2} = E[e^{tY_i^2}] = \frac{1}{(1-2t)^{\frac{1}{2}}} e^{\frac{t\mu_i^2}{1-2t}}, \quad t < \frac{1}{2} \quad (3.2.19)$$

folgt, ergibt sich analog zu Gl. (3.2.13)

$$m_Z(t) = E[e^{tZ}] = \prod_{i=1}^n E[e^{tY_i^2}] = \frac{1}{(1-2t)^{\frac{n}{2}}} e^{\sum_{i=1}^n \frac{t\mu_i^2}{1-2t}} = \frac{1}{(1-2t)^{\frac{n}{2}}} e^{\frac{t\delta}{1-2t}}, \quad t < \frac{1}{2} \quad (3.2.20)$$

Dass die in Gleichung (3.2.16) gegebene Funktion  $f_Z$  die Dichte der  $\chi^2_{n, \delta}$ -Verteilung ist, beweisen wir, indem wir zeigen, dass die durch diese Dichte gegebene Verteilung dieselbe momentenerzeugende Funktion wie  $Z$  hat (s. Satz 1.3.5). Die momentenerzeugende Funktion  $m_{f_Z}$  der durch die Dichte  $f_Z$  gegebenen Verteilung ergibt sich zu

$$\begin{aligned} m_{f_Z}(t) &= \int_0^\infty e^{tz} \sum_{j=0}^\infty \frac{(\frac{\delta}{2})^j}{j!} e^{-\frac{\delta}{2}} \frac{1}{2^{\frac{n+2j}{2}} \Gamma(\frac{n+2j}{2})} z^{(\frac{n+2j}{2}-1)} e^{-\frac{z}{2}} dz \\ &= \sum_{j=0}^\infty \frac{(\frac{\delta}{2})^j}{j!} e^{-\frac{\delta}{2}} \int_0^\infty e^{tz} \frac{1}{2^{\frac{n+2j}{2}} \Gamma(\frac{n+2j}{2})} z^{(\frac{n+2j}{2}-1)} e^{-\frac{z}{2}} dz \end{aligned}$$

Das Integral im  $j$ -ten Summanden ist die momentenerzeugende Funktion der  $\chi^2_{n+2j}$ -Verteilung, hiermit ergibt sich

$$\begin{aligned} m_{f_Z}(t) &= \sum_{j=0}^\infty \frac{(\frac{\delta}{2})^j}{j!} e^{-\frac{\delta}{2}} \frac{1}{(1-2t)^{\frac{n+2j}{2}}} = \frac{1}{(1-2t)^{\frac{n}{2}}} e^{-\frac{\delta}{2}} \sum_{j=0}^\infty \frac{(\frac{\delta}{2(1-2t)})^j}{j!} \\ &= \frac{1}{(1-2t)^{\frac{n}{2}}} e^{-\frac{\delta}{2} + \frac{\delta}{2(1-2t)}} = \frac{1}{(1-2t)^{\frac{n}{2}}} e^{\frac{t\delta}{1-2t}}, \quad t < \frac{1}{2} \end{aligned} \quad (3.2.21)$$

#### Satz 3.2.3. $\chi^2$ -Verteilung

Es sei  $\vec{Y}$  ein  $n$ -dimensionaler Zufallsvektor mit  $\vec{Y} \sim \mathfrak{N}(\vec{\mu}, \Sigma)$  und  $\Sigma$  positiv definit. Dann gilt für die quadratischen Formen  $U$  und  $V$

$$U = (\vec{Y} - \vec{\mu})^T \Sigma^{-1} (\vec{Y} - \vec{\mu}) \sim \chi_n^2 \quad (3.2.22)$$

$$V = \vec{Y}^T \Sigma^{-1} \vec{Y} \sim \chi_{n, \vec{\mu}^T \Sigma^{-1} \vec{\mu}}^2 \quad (3.2.23)$$

**Beweis:**

Da  $\vec{Y} - \vec{\mu} \sim \mathfrak{N}(\vec{0}, \Sigma)$ , ist die Aussage über die Verteilung von  $U$  ein Spezialfall der Aussage über die Verteilung von  $V$ .

Da  $\Sigma$  positiv definit ist, gibt es eine reguläre Matrix  $\Gamma$  mit  $\Sigma = \Gamma\Gamma^T$  (siehe Bemerkung 1 zu Satz 1.3.7). Somit ergibt sich

$$V = \vec{Y}^T \Sigma^{-1} \vec{Y} = \vec{Y}^T (\Gamma\Gamma^T)^{-1} \vec{Y} = \vec{Y}^T (\Gamma^{-1})^T \Gamma^{-1} \vec{Y} = \vec{Z}^T \vec{Z} \quad (3.2.24)$$

mit  $\vec{Z} = \Gamma^{-1} \vec{Y}$ . Da

$$E[\vec{Z}] = \Gamma^{-1} E[\vec{Y}] = \Gamma^{-1} \vec{\mu}, \quad E[\vec{Z}]^T E[\vec{Z}] = \vec{\mu}^T (\Gamma^{-1})^T \Gamma^{-1} \vec{\mu} = \vec{\mu}^T \Sigma^{-1} \vec{\mu} \quad (3.2.25)$$

$$\Sigma_{\vec{Z}} = \Sigma_{\Gamma^{-1} \vec{Y}} = \Gamma^{-1} \Sigma (\Gamma^{-1})^T = \Gamma^{-1} (\Gamma\Gamma^T) (\Gamma^T)^{-1} = I_n \quad (3.2.26)$$

ist nach Satz 3.1.3  $\vec{Z} \sim \mathfrak{N}(\Gamma^{-1} \vec{\mu}, I_n)$ , und somit  $V = \vec{Z}^T \vec{Z} \sim \chi_{n, \vec{\mu}^T \Sigma^{-1} \vec{\mu}}^2$ .

**Definition 3.2.2. t-Verteilung (Student-Verteilung)**

1. Es seien  $U$  und  $V$  unabhängige Zufallsvariablen.  $U$  sei  $\mathfrak{N}(0, 1)$ -verteilt und  $V$   $\chi_n^2$ -verteilt. Die Verteilung von

$$Z = \frac{U}{\sqrt{\frac{1}{n}V}} \quad (3.2.27)$$

heißt *t-Verteilung oder Student-Verteilung mit  $n$  Freiheitsgraden*.

Schreibweise:

$$Z \sim t_n, \quad t_{n;\alpha} := \alpha\text{-Quantil der } t_n\text{-Verteilung}$$

2. Es seien  $U$  und  $V$  unabhängige Zufallsvariablen.  $U$  sei  $\mathfrak{N}(\delta, 1)$ -verteilt und  $V$   $\chi_n^2$ -verteilt. Die Verteilung von

$$Z = \frac{U}{\sqrt{\frac{1}{n}V}} \quad (3.2.28)$$

heißt *nichtzentrale t-Verteilung mit  $n$  Freiheitsgraden und dem Nichtzentralitätsparameter  $\delta$* .

Schreibweise:

$$Z \sim t_{n,\delta}, \quad t_{n,\delta;\alpha} := \alpha\text{-Quantil der } t_{n,\delta}\text{-Verteilung}$$

**Satz 3.2.4. t-Verteilung (Student-Verteilung)**

1. Es sei  $Z$  eine  $t_n$ -verteilte Zufallsgröße, dann gilt:

$$f_Z(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \frac{1}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}}, \quad t \in \mathbb{R} \quad (3.2.29)$$

$$f_{t_n}(t) = f_{t_n}(-t) \quad (\text{Symmetrie zu } 0) \quad (3.2.30)$$

$$t_{n;\alpha} = -t_{n;1-\alpha} \quad (3.2.31)$$

$$E[Z] = 0 \quad \text{für } n > 1 \quad (3.2.32)$$

$$\text{Var}(Z) = \frac{n}{n-2} \quad \text{für } n > 2 \quad (3.2.33)$$



2. Für die Verteilungsfunktion  $F_{t_n}$  der  $t_n$ -Verteilung gilt:

$$\lim_{n \rightarrow \infty} F_{t_n}(t) = \Phi(t) \quad (3.2.34)$$

### Definition 3.2.3. F-Verteilung

1. Es seien  $V_1$  und  $V_2$  unabhängige  $\chi^2$ -verteilte Zufallsgrößen mit  $n_1$  bzw.  $n_2$  Freiheitsgraden. Dann heißt die Verteilung von

$$Z = \frac{\frac{1}{n_1} V_1}{\frac{1}{n_2} V_2} \quad (3.2.35)$$

$F$ -Verteilung mit  $n_1$  und  $n_2$  Freiheitsgraden.

Schreibweise:

$$Z \sim F_{n_1, n_2}; \quad F_{n_1, n_2; \alpha} := \alpha\text{-Quantil der } F_{n_1, n_2}\text{-Verteilung}$$

2. Es seien  $V_1$  und  $V_2$  unabhängige Zufallsgrößen mit  $V_1 \sim \chi_{n_1, \delta}^2$  und  $V_2 \sim \chi_{n_2}^2$ . Dann heißt die Verteilung von

$$Z = \frac{\frac{1}{n_1} V_1}{\frac{1}{n_2} V_2} \quad (3.2.36)$$

nichtzentrale  $F$ -Verteilung mit  $n_1$  und  $n_2$  Freiheitsgraden und Nichtzentralitätsparameter  $\delta$ .

Schreibweise:

$$Z \sim F_{n_1, n_2, \delta}; \quad F_{n_1, n_2, \delta; \alpha} := \alpha\text{-Quantil der } F_{n_1, n_2, \delta}\text{-Verteilung}$$

### Satz 3.2.5. F-Verteilung

1. Es sei  $Z$  eine  $F$ -verteilte Zufallsgröße mit  $n_1$  und  $n_2$  Freiheitsgraden, dann gilt:

$$f_Z(z) = \begin{cases} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right) \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{z^{\frac{n_1-2}{2}}}{\left(1+\frac{n_1}{n_2}z\right)^{\frac{n_1+n_2}{2}}} & \text{für } z > 0 \\ 0 & \text{für } z \leq 0 \end{cases} \quad (3.2.37)$$

$$E[Z] = \frac{n_2}{n_2 - 2}, \quad n_2 > 2 \quad (3.2.38)$$

$$\text{Var}(Z) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4 \quad (3.2.39)$$

2. Es sei  $Z_n$   $F_{m, n}$ -verteilt. Dann ist

$$U_n = mZ_n \quad (3.2.40)$$

asymptotisch  $\chi_m^2$ -verteilt. Für die Verteilungsfunktion der  $U_n$  gilt:

$$\lim_{n \rightarrow \infty} F_{U_n}(u) = F_{\chi_m^2}(u) \quad (3.2.41)$$

### 3.3 Die Maximum-Likelihood-Methode und die Verteilung der Schätzer

#### Definition 3.3.1. Klassisches lineares Modell

Das klassische lineare Modell ist charakterisiert durch die Gleichungen und die Verteilungsannahme

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^k \times \mathbb{R}_{>0} \quad (3.3.1)$$

Dabei ist  $\vec{Y} = (Y_1, \dots, Y_n)^T$  der Zufallsvektor der beobachteten Größen,  $\vec{\beta} = (\beta_1, \dots, \beta_k)^T$  der Vektor der unbekannten konstanten Parameter,  $X = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, k}}$  die konstante Designmatrix,  $\vec{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  der Zufallsvektor der Fehler,  $\sigma^2 > 0$  die Varianz der Fehlergrößen,  $I$  die  $(n \times n)$ -Einheitsmatrix und  $\Omega$  der Parameterraum des Modells.

#### Bemerkung 1.

Für das durch Gleichung (3.3.1) gegebene klassische lineare Modell gilt (s. Satz 3.1.3)

$$\vec{Y} \sim \mathfrak{N}(X\vec{\beta}, \sigma^2 I). \quad (3.3.2)$$

Die Fehlergrößen  $\epsilon_1, \dots, \epsilon_n$  sowie die Beobachtungsgrößen  $Y_1, \dots, Y_n$  sind unabhängige normalverteilte Zufallsgrößen mit der Varianz  $\sigma^2$ .

#### Bemerkung 2.

Durch (3.3.1) ist ein parametrisches Modell für die Verteilung des Stichprobenvektors  $\vec{Y}$  gegeben. Somit kann die Maximum-Likelihood-Methode zur Gewinnung von Schätzern für die unbekannten Modellparameter  $\vec{\beta}$  und  $\sigma^2$  verwendet werden.

#### Satz 3.3.1. Maximum-Likelihood-Schätzung

Für das durch Gleichung (3.3.1) gegebene klassische lineare Modell sind die Maximum-Likelihood-Schätzwerte für  $\vec{\beta}$  genau die Lösungen  $\hat{\beta}$  der Minimierungsaufgabe

$$\|\vec{Y} - X\hat{\beta}\| = \min_{\vec{\beta} \in \mathbb{R}^k} \|\vec{Y} - X\vec{\beta}\|, \quad (3.3.3)$$

und somit der Normalgleichungen (s. Satz 2.1.1)

$$X^T X \hat{\beta} = X^T \vec{Y}. \quad (3.3.4)$$

Die Maximum-Likelihood-Methode und die Methode der Kleinsten Quadrate liefern für den Parametervektor  $\vec{\beta}$  dieselben Ergebnisse.

Die Maximum-Likelihood-Schätzung für  $\sigma^2$  ist

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} (\vec{Y} - X\hat{\beta})^T (\vec{Y} - X\hat{\beta}). \quad (3.3.5)$$

Sie unterscheidet sich von der in Gleichung (2.2.39) eingeführten erwartungstreuen Schätzung  $\hat{\sigma}^2$  nur durch einen konstanten Proportionalitätsfaktor, denn es gilt

$$\hat{\sigma}_{ML}^2 = \frac{n-r}{n} \hat{\sigma}^2 \quad (3.3.6)$$

**Beweis**

Da  $\vec{Y} \sim \mathfrak{N}(X\vec{\beta}, \sigma^2 I)$  ergibt sich die Dichte von  $\vec{Y}$  zu (s. Satz 3.1.6)

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2}(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta})} \quad (3.3.7)$$

Somit lautet die Likelihoodfunktion, die Dichte aufgefasset als Funktion der unbekannten Parameter,

$$L(\vec{\beta}, \sigma^2; \vec{y}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2}(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta})} \quad (3.3.8)$$

und die Loglikelihoodfunktion

$$\log L(\vec{\beta}, \sigma^2; \vec{y}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2}(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta}) \quad (3.3.9)$$

Die Maximum-Likelihood-Schätzwerte  $\hat{\beta}_{ML}$  und  $\hat{\sigma}_{ML}^2$  für  $\vec{\beta}$  und  $\sigma^2$  ergeben sich durch Maximierung der Likelihoodfunktion bzw. der Loglikelihoodfunktion

$$\log L(\hat{\beta}_{ML}, \hat{\sigma}_{ML}^2; \vec{y}) = \max_{\vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0} \log L(\vec{\beta}, \sigma^2; \vec{y}), \quad (3.3.10)$$

Die Maximierung kann durchgeführt werden, indem zunächst für einen beliebigen, jedoch fest vorgegebenen Wert von  $\sigma^2$  bezüglich  $\vec{\beta}$  und anschließend bezüglich  $\sigma^2$  maximiert wird. Für einen fest vorgegebenen Wert von  $\sigma^2$  wird die Loglikelihoodfunktion bezüglich  $\vec{\beta}$  genau dann maximiert, wenn

$$(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta}) \quad (3.3.11)$$

bezüglich  $\vec{\beta}$  minimiert wird. Es sei  $\tilde{\beta}$  eine beliebige Lösung dieser Minimierungsaufgabe

$$\|\vec{y}-X\tilde{\beta}\| = \min_{\vec{\beta} \in \mathbb{R}^k} \|\vec{y}-X\vec{\beta}\|, \quad (3.3.12)$$

die unabhängig von dem konkreten Wert von  $\sigma^2$  ist, dann ergibt sich

$$\begin{aligned} g(\sigma^2) &= \max_{\vec{\beta} \in \mathbb{R}^k} \log L(\vec{\beta}, \sigma^2; \vec{y}) = \log L(\tilde{\beta}, \sigma^2; \vec{y}) \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2}(\vec{y}-X\tilde{\beta})^T(\vec{y}-X\tilde{\beta}) \end{aligned} \quad (3.3.13)$$

Differentiation dieser Funktion nach  $\sigma^2$  und Nullsetzen der Ableitung ergibt die Gleichung

$$\frac{d}{d\sigma^2} g(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{(\vec{y}-X\tilde{\beta})^T(\vec{y}-X\tilde{\beta})}{2\sigma^4} = 0 \quad (3.3.14)$$

mit der Lösung

$$\tilde{\sigma}^2 = \frac{1}{n}(\vec{y}-X\tilde{\beta})^T(\vec{y}-X\tilde{\beta}) \quad (3.3.15)$$

Da sich für die zweite Ableitung

$$\frac{d^2}{(d\sigma^2)^2} g(\sigma^2) \Big|_{\sigma^2=\tilde{\sigma}^2} = \frac{n}{2\tilde{\sigma}^4} - \frac{\|\vec{y}-X\tilde{\beta}\|^2}{\tilde{\sigma}^6} = \frac{n}{2\tilde{\sigma}^4} - \frac{n}{\tilde{\sigma}^4} = -\frac{n}{2\tilde{\sigma}^4} < 0 \quad (3.3.16)$$

ergibt, ist  $\tilde{\sigma}^2$  Maximalstelle.

**Satz 3.3.2. Verteilung der Schätzer**

Sei  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  eine  $q$ -dimensionale schätzbare Funktion für das durch Gleichung (3.3.1) gegebene klassische lineare Modell. Dann gelten für den BLE-Schätzer  $\hat{\psi} = C\hat{\beta} = A\vec{Y}$  und die Schätzung der Varianz  $\hat{\sigma}^2 = \frac{1}{n-r}(\vec{Y} - X\hat{\beta})^T(\vec{Y} - X\hat{\beta})$  bei beliebigem  $(\vec{\beta}^T, \sigma^2)^T \in \Omega$  die folgenden Aussagen, hierbei ist  $\hat{\beta}$  eine beliebige Lösung der Normalgleichungen und die Zeilenvektoren der eindeutig bestimmten Matrix  $A$  liegen im Spaltenraum von  $X$ :

1.  $\hat{\psi} \sim \mathfrak{N}(\vec{\psi}(\vec{\beta}), \sigma^2 C(X^T X)^{-1} C^T = \sigma^2 A A^T)$
2.  $\frac{n-r}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} \hat{\epsilon}^T \hat{\epsilon} = \frac{1}{\sigma^2} (\vec{Y} - X\hat{\beta})^T (\vec{Y} - X\hat{\beta}) \sim \chi_{n-r}^2$
3.  $\hat{\psi}$  und  $\hat{\sigma}^2$  sind unabhängig.

**Beweis:**

Beweis von 1:

Da  $\vec{Y} \sim \mathfrak{N}(X\vec{\beta}, \sigma^2 I)$ , ist  $\hat{\psi} = A\vec{Y}$  normalverteilt (s. Satz 3.1.3).  $E[\hat{\psi}] = \vec{\psi}(\vec{\beta})$  folgt aus der Erwartungstreue von  $\hat{\psi}$ . Ferner gilt  $\Sigma_{\hat{\psi}} = \Sigma_{A\vec{Y}} = A\Sigma_{\vec{Y}}A^T = \sigma^2 A A^T$ . Aus Satz 2.1.6 folgt  $\Sigma_{\hat{\psi}} = \sigma^2 C(X^T X)^{-1} C^T$ .

Beweis von 2 und 3:

Zum Beweis verwenden wir eine kanonische Darstellung des linearen Modells. Es sei  $Q$  eine Matrix, deren Spaltenvektoren eine Orthonormalbasis des  $\mathbb{R}^n$  bilden, wobei die ersten  $r$  Spaltenvektoren den Spaltenraum von  $X$  aufspannen. Es sei  $\vec{Z} = Q^T \vec{Y}$  und

$$Q_1 = \begin{bmatrix} \vec{q}_1 & \dots & \vec{q}_r \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \vec{q}_{r+1} & \dots & \vec{q}_n \end{bmatrix}$$

$$\vec{Z}_1 = (Z_1, \dots, Z_r)^T, \quad \vec{Z}_2 = (Z_{r+1}, \dots, Z_n)^T \quad (3.3.17)$$

Aus der Normalverteilung von  $\vec{Y}$  und den in Satz 2.2.2 hergeleiteten Eigenschaften der kanonischen Darstellung folgt:

$$\vec{Z} \sim \mathfrak{N}(E[\vec{Z}], \sigma^2 I)$$

$$E[\vec{Z}] = (\gamma_1, \dots, \gamma_r, 0, \dots, 0)^T$$

$$\hat{\epsilon}^T \hat{\epsilon} = \sum_{i=r+1}^n Z_i^2 = \vec{Z}_2^T \vec{Z}_2 \quad (3.3.18)$$

hiermit ergibt sich

$$\frac{1}{\sigma} \vec{Z} \sim \mathfrak{N}\left(\frac{1}{\sigma} E[\vec{Z}], I\right) \quad (3.3.19)$$

Somit sind die Größen  $\frac{1}{\sigma} Z_{r+1}, \dots, \frac{1}{\sigma} Z_n$  unabhängig und standardisiert normalverteilt und die Zufallsvektoren  $\vec{Z}_1$  und  $\vec{Z}_2$  stochastisch unabhängig (s. Satz 3.1.3 und Satz 3.1.4). Hieraus folgt

$$\frac{1}{\sigma^2} \hat{\epsilon}^T \hat{\epsilon} = \sum_{i=r+1}^n \left( \frac{Z_i}{\sigma} \right)^2 = \frac{1}{\sigma^2} \vec{Z}_2^T \vec{Z}_2 \sim \chi_{n-r}^2 \quad (3.3.20)$$

Da die Zeilenvektoren von  $A$  im Spaltenraum von  $X$  und somit im Spaltenraum von  $Q_1$  liegen, ergibt sich:

$$\hat{\psi} = A\vec{Y} = AQ\vec{Z} = A[Q_1 \ Q_2] \begin{bmatrix} \vec{Z}_1 \\ \vec{Z}_2 \end{bmatrix} = [AQ_1 \ O] \begin{bmatrix} \vec{Z}_1 \\ \vec{Z}_2 \end{bmatrix} = AQ_1\vec{Z}_1 \quad (3.3.21)$$

Somit ist  $\hat{\psi}$  eine Funktion von  $\vec{Z}_1$  und  $\hat{\sigma}^2$  eine Funktion von  $\vec{Z}_2$ . Aus der Unabhängigkeit von  $\vec{Z}_1$  und  $\vec{Z}_2$  folgt die Unabhängigkeit von  $\hat{\psi}$  und  $\hat{\sigma}^2$  (s. Satz 1.3.2(1)).

**Korollar 1.**

*Die Designmatrix des durch Gleichung (3.3.1) gegebenen klassischen linearen Modells besitze vollen Spaltenrang. Dann gelten für den BLE-Schätzer  $\hat{\beta}$  für  $\vec{\beta}$  und die Schätzung  $\hat{\sigma}^2$  für  $\sigma^2$ :*

1.  $\hat{\beta} = (X^T X)^{-1} X^T \vec{Y} \sim \mathfrak{N}(\vec{\beta}, \sigma^2 (X^T X)^{-1})$
2.  $\frac{n-k}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} \hat{\epsilon}^T \hat{\epsilon} = \frac{1}{\sigma^2} (\vec{Y} - X\hat{\beta})^T (\vec{Y} - X\hat{\beta}) \sim \chi_{n-k}^2$
3.  $\hat{\beta}$  und  $\hat{\sigma}^2$  sind unabhängig.



## Kapitel 4

# Konfidenzbereiche

### 4.1 Konfidenzintervalle und Vorhersageintervalle

#### Definition 4.1.1. Konfidenzbereich für eine schätzbare Funktion

Es sei  $\vec{\psi}(\vec{\beta}) = (\psi_1(\vec{\beta}), \dots, \psi_q(\vec{\beta}))^T = C\vec{\beta}$  eine  $q$ -dimensionale schätzbare Funktion des Parametervektors  $\vec{\beta}$  des klassischen linearen Modells für den  $n$ -dimensionalen Beobachtungsvektor  $\vec{Y}$

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^k \times \mathbb{R}_{>0} \quad (4.1.1)$$

Ein Konfidenzbereich für  $\vec{\psi}$  bzw. ein gemeinsamer Konfidenzbereich für  $\psi_1, \dots, \psi_q$  zum Niveau  $(1 - \alpha)$  ( $0 < \alpha < 1$ ) ist eine Abbildung  $K$  auf dem Stichprobenraum  $(\mathbb{R}^n)$ , deren Bilder Teilmengen des  $\mathbb{R}^q$  sind, so dass gilt:

$$P_{\vec{\beta}, \sigma^2}(\vec{\psi}(\vec{\beta}) \in K(\vec{Y})) \geq 1 - \alpha \quad \forall \vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0 \quad (4.1.2)$$

#### Satz 4.1.1. Konfidenzintervall für eine eindimensionale schätzbare Funktion

Es sei  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  mit  $\vec{c} \neq \vec{0}$  eine schätzbare Funktion des Parametervektors  $\vec{\beta}$  des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells,  $\hat{\psi} = \vec{c}^T \hat{\beta} = \vec{a}^T \vec{Y}$  der BLE-Schätzer von  $\psi$  und

$$\hat{\sigma}^2 = \frac{1}{n - r} (\vec{Y} - X\hat{\beta})^T (\vec{Y} - X\hat{\beta}) \quad (4.1.3)$$

die erwartungstreue Schätzung der Varianz  $\sigma^2$ . Hierbei ist  $\hat{\beta}$  eine beliebige Lösung der Normalgleichungen und der Vektor  $\vec{a}$  liegt im Spaltenraum von  $X$ . Für einen beliebige Vektor  $(\vec{\beta}^T, \sigma^2)^T \in \Omega$  gelten die folgenden Aussagen:

1.  $\hat{\psi} \sim \mathfrak{N}(\psi(\vec{\beta}), \sigma_{\hat{\psi}}^2)$ ,  $\sigma_{\hat{\psi}}^2 = \sigma^2 \vec{c}^T (X^T X)^{-} \vec{c} = \sigma^2 \vec{a}^T \vec{a}$
2.  $\frac{\hat{\psi} - \psi(\vec{\beta})}{\sigma_{\hat{\psi}}} \sim \mathfrak{N}(0, 1)$
3.  $\hat{\sigma}_{\hat{\psi}}^2 = \hat{\sigma}^2 \vec{a}^T \vec{a}$  ist eine erwartungstreue Schätzung der Varianz von  $\hat{\psi}$ .

$$4. \quad (n-r) \frac{\hat{\sigma}_{\hat{\psi}}^2}{\sigma_{\hat{\psi}}^2} = (n-r) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2$$

5.  $\hat{\psi}$  und  $\hat{\sigma}_{\hat{\psi}}^2$  sind unabhängig.

$$6. \quad \frac{\hat{\psi} - \psi(\vec{\beta})}{\hat{\sigma}_{\hat{\psi}}} = \frac{\frac{\hat{\psi} - \psi(\vec{\beta})}{\sigma_{\hat{\psi}}}}{\sqrt{\frac{\hat{\sigma}_{\hat{\psi}}^2}{\sigma_{\hat{\psi}}^2}}} \sim t_{n-r}$$

$$7. \quad P_{\vec{\beta}, \sigma^2} \left( -t_{n-r; 1-\frac{\alpha}{2}} \leq \frac{\hat{\psi} - \psi(\vec{\beta})}{\hat{\sigma}_{\hat{\psi}}} \leq t_{n-r; 1-\frac{\alpha}{2}} \right) = 1 - \alpha \quad \forall \vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0$$

8. Das zufällige Intervall

$$\begin{aligned} K(\vec{Y}) &= \left\{ \psi \in \mathbb{R} \mid -t_{n-r; 1-\frac{\alpha}{2}} \leq \frac{\hat{\psi} - \psi}{\hat{\sigma}_{\hat{\psi}}} \leq t_{n-r; 1-\frac{\alpha}{2}} \right\} \\ &= \left\{ \psi \in \mathbb{R} \mid \hat{\psi} - t_{n-r; 1-\frac{\alpha}{2}} \hat{\sigma}_{\hat{\psi}} \leq \psi \leq \hat{\psi} + t_{n-r; 1-\frac{\alpha}{2}} \hat{\sigma}_{\hat{\psi}} \right\} \\ &= [\hat{\psi} - t_{n-r; 1-\frac{\alpha}{2}} \hat{\sigma}_{\hat{\psi}}, \hat{\psi} + t_{n-r; 1-\frac{\alpha}{2}} \hat{\sigma}_{\hat{\psi}}] \end{aligned} \quad (4.1.4)$$

ist ein  $(1 - \alpha)$ -Konfidenzintervall für  $\psi$ .

#### Beweis:

Die Aussagen des Satzes folgen direkt aus Satz 3.3.2 und der Definition der  $t$ -Verteilung (Def. 3.2.2).

#### Beispiel 4.1.1. Konfidenzintervall eines Parameters bei vollem Spaltenrang

Die Designmatrix des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells besitze vollen Spaltenrang, d.h.  $\text{rg}(X) = r = k < n$ . Es sei  $\hat{\beta}$  der BLE-Schätzer von  $\vec{\beta}$  und  $\hat{\sigma}^2$  die durch Gleichung (4.1.3) gegebene erwartungstreue Schätzung für  $\sigma^2$ . Dann ist  $\beta_i$ , die  $i$ -te Komponente des Parametervektors  $\vec{\beta}$ , schätzbar. Aus Satz 4.1.1 folgt mit  $\vec{c} = \vec{e}_i$  ( $\vec{e}_i$  sei der  $i$ -te Einheitsvektor) für den BLE-Schätzer  $\hat{\psi} = \hat{\beta}_i$  von  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta} = \beta_i$ :

$$1. \quad \hat{\psi} = \hat{\beta}_i \sim \mathcal{N}(\beta_i, \sigma^2 d_i), \quad d_i \text{ ist das } i\text{-te Diagonalelement von } (X^T X)^{-1}.$$

$$2. \quad \hat{\sigma}_{\hat{\psi}}^2 = \hat{\sigma}_{\hat{\beta}_i}^2 = \hat{\sigma}^2 d_i$$

$$3. \quad \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma} \sqrt{d_i}} \sim t_{n-k}$$

4. Das zufällige Intervall

$$K(\vec{Y}) = [\hat{\beta}_i - t_{n-k; 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{d_i}, \hat{\beta}_i + t_{n-k; 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{d_i}] \quad (4.1.5)$$

ist ein  $(1 - \alpha)$ -Konfidenzintervall für  $\beta_i$ .



**Beispiel 4.1.2. Einstichprobenproblem**

Die Aufgabe, aufgrund einer Zufallsstichprobe vom Umfang  $n$  aus einer  $\mathfrak{N}(\mu, \sigma^2)$ -verteilten Grundgesamtheit ein Konfidenzintervall für  $\mu$  anzugeben, kann in das klassische lineare Modell

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^k \times \mathbb{R}_{>0} \quad (4.1.6)$$

wie folgt eingebettet werden. Es ist  $\vec{Y} = (Y_1, \dots, Y_n)^T$  der Vektor der Stichprobenvariablen,  $X = \vec{1}_n$ ,  $\text{rg}(X) = k = 1 < n$  und  $\vec{\beta} = \mu$ . Hieraus folgt

$$\hat{\mu} = (X^T X)^{-1} X^T \vec{Y} = (\vec{1}_n^T \vec{1}_n)^{-1} \vec{1}_n^T \vec{Y} = \frac{1}{n} \vec{1}_n^T \vec{Y} = \bar{Y} \quad (4.1.7)$$

$$(X^T X)^{-1} = (\vec{1}_n^T \vec{1}_n)^{-1} = \frac{1}{n}, \quad \text{Var}(\hat{\mu}) = \sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{n} \quad (4.1.8)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \hat{\epsilon}^T \hat{\epsilon} = \frac{1}{n-1} (\vec{Y} - \vec{1}_n \bar{Y})^T (\vec{Y} - \vec{1}_n \bar{Y}) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (4.1.9)$$

Mit Satz 4.1.1 und Beispiel 4.1.1 ergeben sich die Verteilungsaussagen

1.  $\hat{\mu} \sim \mathfrak{N}(\mu, \frac{\sigma^2}{n})$
2.  $\frac{n-1}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-1}^2$
3.  $\hat{\mu}$  und  $\hat{\sigma}^2$  sind unabhängig.
4.  $\sqrt{n} \frac{\hat{\mu} - \mu}{\hat{\sigma}} \sim t_{n-1}$

und für  $\mu$  das  $(1 - \alpha)$ -Konfidenzintervall

$$K(\vec{Y}) = \left[ \hat{\mu} - t_{n-1; 1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + t_{n-1; 1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} \right] \quad (4.1.10)$$

**Beispiel 4.1.3. Konfidenzintervall des Erwartungswertes einer Beobachtung**

Es sei  $\vec{Y}$  der Beobachtungsvektor des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells. Gesucht ist ein Konfidenzintervall basierend auf  $\vec{Y}$  für den Erwartungswert  $E[Y] = y(\vec{x}) = \vec{x}^T \vec{\beta}$  ( $\vec{x} \in \mathfrak{I}\mathfrak{m}(X^T)$ ) einer Beobachtung  $Y$  des linearen Modells zu vorgegebenen Werten der Einflussgrößen:

$$Y = y(\vec{x}) + \epsilon = \vec{x}^T \vec{\beta} + \epsilon, \quad \epsilon \sim \mathfrak{N}(0, \sigma^2), \quad \vec{x} \in \mathfrak{I}\mathfrak{m}(X^T) \quad (4.1.11)$$

Die Parameterfunktion  $\psi(\vec{\beta}) = y(\vec{x}) = \vec{x}^T \vec{\beta}$  ist schätzbar, da  $\vec{x} \in \mathfrak{I}\mathfrak{m}(X^T)$ , mit dem BLE-Schätzer  $\hat{\psi} = \hat{y}(\vec{x}) = \vec{x}^T \hat{\beta}$ . Somit ergibt Satz 4.1.1 mit  $\vec{c} = \vec{x}$  für  $y(\vec{x})$  das  $(1 - \alpha)$ -Konfidenzintervall

$$K(\vec{Y}) = [\hat{y}(\vec{x}) - t_{n-r; 1-\frac{\alpha}{2}} \hat{\sigma} h(\vec{x}), \hat{y}(\vec{x}) + t_{n-r; 1-\frac{\alpha}{2}} \hat{\sigma} h(\vec{x})] \quad (4.1.12)$$

wobei

$$h(\vec{x}) = \sqrt{\vec{x}^T (X^T X)^{-1} \vec{x}} \quad (4.1.13)$$

**Definition 4.1.2. Vorhersageintervall für eine Beobachtung**

Es seien  $Y, Y_1, \dots, Y_n$  Zufallsgrößen mit einer gemeinsamen Verteilung aus einer gegebenen Verteilungsfamilie. Ein Vorhersageintervall zum Niveau  $(1 - \alpha)$  ( $0 < \alpha < 1$ ) von  $Y$  basierend auf  $Y_1, \dots, Y_n$  ist eine Abbildung  $V$  auf dem Stichprobenraum  $(\mathbb{R}^n)$  mit Intervallen als Bilder, so dass für jede Verteilung der gegebenen Verteilungsfamilie gilt:

$$P(Y \in V(Y_1, \dots, Y_n)) \geq 1 - \alpha \quad (4.1.14)$$

**Beispiel 4.1.4. Vorhersageintervall für eine Beobachtung**

Es sei  $\vec{Y}$  der Beobachtungsvektor des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells. Gesucht ist ein Vorhersageintervall basierend auf  $\vec{Y}$  für eine zukünftige Beobachtung  $Y$  des linearen Modells zu vorgegebenen Werten der Einflussgrößen.

Es sei

$$Y = y(\vec{x}) + \epsilon = \vec{x}^T \vec{\beta} + \epsilon, \quad \epsilon \sim \mathfrak{N}(0, \sigma^2), \quad \vec{x} \in \mathfrak{I}(\mathbf{m}(X^T)) \quad (4.1.15)$$

Die Beobachtung  $Y$  sei unabhängig von den Beobachtungen  $Y_1, \dots, Y_n$  des zugrundeliegenden klassischen linearen Modells, und  $\hat{y}(\vec{x}) = \vec{x}^T \hat{\beta}$  sei der BLE-Schätzer des Erwartungswert von  $Y$ . Für  $Y - \hat{y}(\vec{x})$  ergibt sich:

1.  $E[Y - \hat{y}(\vec{x})] = E[Y - \vec{x}^T \hat{\beta}] = E[Y] - E[\vec{x}^T \hat{\beta}] = \vec{x}^T \vec{\beta} - \vec{x}^T \vec{\beta} = 0$
2. Die unabhängigen Zufallsgrößen  $\vec{x}^T \hat{\beta}$  und  $\hat{\sigma}^2$  sind Funktionen von  $\vec{Y}$ . Da  $Y$  und  $\vec{Y}$  unabhängig sind, sind auch  $Y$  und  $(\vec{x}^T \hat{\beta}, \hat{\sigma}^2)$  unabhängig. Somit sind auch die Zufallsgrößen  $Y$ ,  $\vec{x}^T \hat{\beta}$  und  $\hat{\sigma}^2$  unabhängig.
3. Aus Satz 4.1.1 und Gleichung (4.1.15) folgt

$$\begin{aligned} \text{Var}(Y - \hat{y}(\vec{x})) &= \text{Var}(Y - \vec{x}^T \hat{\beta}) = \text{Var}(Y) + \text{Var}(\vec{x}^T \hat{\beta}) \\ &= \sigma^2 + \sigma^2 \vec{x}^T (X^T X)^{-1} \vec{x} \\ Y - \hat{y}(\vec{x}) &= Y - \vec{x}^T \hat{\beta} \sim \mathfrak{N}(0, \sigma^2(1 + \vec{x}^T (X^T X)^{-1} \vec{x})) \end{aligned} \quad (4.1.16)$$

4. Somit ergibt sich

$$\frac{Y - \hat{y}(\vec{x})}{\hat{\sigma} \sqrt{1 + \vec{x}^T (X^T X)^{-1} \vec{x}}} = \frac{\frac{Y - \hat{y}(\vec{x})}{\sigma \sqrt{1 + \vec{x}^T (X^T X)^{-1} \vec{x}}}}{\frac{\hat{\sigma}}{\sigma}} \sim t_{n-r} \quad (4.1.17)$$

hieraus folgt

$$P_{\vec{\beta}, \sigma^2} \left( -t_{n-r; 1-\frac{\alpha}{2}} \leq \frac{Y - \hat{y}(\vec{x})}{\hat{\sigma} \sqrt{1 + \vec{x}^T (X^T X)^{-1} \vec{x}}} \leq t_{n-r; 1-\frac{\alpha}{2}} \right) = 1 - \alpha \quad \forall \vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0 \quad (4.1.18)$$

Somit ist

$$V(\vec{Y}) = \left[ \hat{y}(\vec{x}) - t_{n-r;1-\frac{\alpha}{2}} \hat{\sigma} h(\vec{x}), \hat{y}(\vec{x}) + t_{n-r;1-\frac{\alpha}{2}} \hat{\sigma} h(\vec{x}) \right] \quad (4.1.19)$$

wobei

$$h(\vec{x}) = \sqrt{1 + \vec{x}^T (X^T X)^{-1} \vec{x}} \quad (4.1.20)$$

ein Vorhersageintervall zum Niveau  $1 - \alpha$  für  $Y$  basierend auf  $\vec{Y}$ .

#### Beispiel 4.1.5. Konfidenz- und Vorhersageintervalle für die einfache lineare Regression

Für das durch die Gleichungen

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (i = 1, \dots, n), \quad \sigma^2 > 0 \quad (4.1.21)$$

$\epsilon_1, \dots, \epsilon_n$  stochastisch unabhängig

gegebene einfache lineare Regressionsmodell ergeben sich aus dem Korollar 1 zu Satz 3.3.2, den Gleichungen (2.1.38) und dem Beispiel 4.1.1 die folgenden Verteilungsaussagen für die BLE-Schätzer  $\hat{\beta}_0$  für  $\beta_0$ ,  $\hat{\beta}_1$  für  $\beta_1$  und die Schätzung  $\hat{\sigma}^2$  für  $\sigma^2$ :

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \frac{\sigma^2}{n s_{xx}} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \right)$$

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \sim \chi_{n-2}^2$$

$(\hat{\beta}_0, \hat{\beta}_1)^T$  und  $\hat{\sigma}^2$  sind unabhängig. (4.1.22)

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma} \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{n s_{xx}}}} \sim t_{n-2}, \quad \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma} \sqrt{\frac{1}{n s_{xx}}}} \sim t_{n-2}$$

sowie die  $(1 - \alpha)$ -Konfidenzintervalle

$$K_{\beta_0}(\vec{Y}) = \left[ \hat{\beta}_0 - t_{n-2;1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{n s_{xx}}}, \hat{\beta}_0 + t_{n-2;1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{n s_{xx}}} \right]$$

$$K_{\beta_1}(\vec{Y}) = \left[ \hat{\beta}_1 - t_{n-2;1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{1}{n s_{xx}}}, \hat{\beta}_1 + t_{n-2;1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{1}{n s_{xx}}} \right] \quad (4.1.23)$$

Für die alternative Parametrisierung

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (i = 1, \dots, n), \quad \sigma^2 > 0 \quad (4.1.24)$$

$\epsilon_1, \dots, \epsilon_n$  stochastisch unabhängig

ergibt sich analog mit den Gleichungen (2.1.45) für die BLE-Schätzer  $\hat{\alpha}$  für  $\alpha$  und  $\hat{\beta}$  für  $\beta$ :

$$\begin{aligned} \hat{\alpha} &\sim \mathfrak{N}\left(\alpha, \frac{\sigma^2}{n}\right), \quad \hat{\beta} \sim \mathfrak{N}\left(\beta, \frac{\sigma^2}{ns_{xx}}\right), \quad \hat{\alpha} \text{ und } \hat{\beta} \text{ sind unabhängig.} \\ \frac{n-2}{\sigma^2} \hat{\sigma}^2 &= \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - (\hat{\alpha} + \hat{\beta}(x_i - \bar{x})))^2 \sim \chi_{n-2}^2 \\ \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2 &\text{ sind unabhängig.} \\ \frac{\hat{\alpha} - \alpha}{\frac{\hat{\sigma}}{\sqrt{n}}} &\sim t_{n-2}, \quad \frac{\hat{\beta} - \beta}{\hat{\sigma} \sqrt{\frac{1}{ns_{xx}}}} \sim t_{n-2} \end{aligned} \quad (4.1.25)$$

sowie die  $(1 - \alpha)$ -Konfidenzintervalle

$$\begin{aligned} K_{\alpha}(\vec{Y}) &= \left[ \hat{\alpha} - t_{n-2; 1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\alpha} + t_{n-2; 1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} \right] \\ K_{\beta}(\vec{Y}) &= K_{\beta_1}(\vec{Y}) \end{aligned} \quad (4.1.26)$$

Für die Schätzung  $\hat{y}(x)$  der Regressionsgeraden

$$y(x) = \beta_0 + \beta_1 x = \alpha + \beta(x - \bar{x}) \quad (4.1.27)$$

an der Stelle  $x$  ergibt sich mit Gleichung (2.1.48)

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = \hat{\alpha} + \hat{\beta}(x - \bar{x}) \sim \mathfrak{N}\left(y(x), \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}} \right)\right) \quad (4.1.28)$$

Da  $(\hat{\beta}_0, \hat{\beta}_1)^T$  und  $\hat{\sigma}^2$  unabhängig sind, sind auch  $\hat{y}(x)$  und  $\hat{\sigma}^2$  unabhängig, und es ergibt sich

$$\frac{\hat{y}(x) - y(x)}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}}}} \sim t_{n-2} \quad (4.1.29)$$

Hieraus folgt, dass

$$K_{y(x)}(\vec{Y}) = \left[ \hat{y}(x) - t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}}}, \hat{y}(x) + t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}}} \right] \quad (4.1.30)$$

ein  $(1 - \alpha)$ -Konfidenzintervall für  $y(x) = \beta_0 + \beta_1 x$  darstellt.

Ein Vorhersageintervall für eine zukünftige Beobachtung für einen vorgegebenen Wert der Einflussgröße  $x$  ergibt sich analog zu Beispiel 4.1.4. Es sei

$$Y = y(x) + \epsilon = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim \mathfrak{N}(0, \sigma^2) \quad (4.1.31)$$

Die Beobachtung  $Y$  sei unabhängig von den Beobachtungen  $Y_1, \dots, Y_n$  des zugrundeliegenden klassischen linearen Modells. Für  $Y - \hat{y}(x)$  ergibt sich:

1.  $E[Y - \hat{y}(x)] = E[Y - (\hat{\beta}_0 + \hat{\beta}_1 x)] = E[Y] - (E[\hat{\beta}_0] + E[\hat{\beta}_1]x)$   
 $= \beta_0 + \beta_1 x - (\beta_0 + \beta_1 x) = 0$
2. Die unabhängigen Zufallsgrößen  $(\hat{\beta}_0, \hat{\beta}_1)^T$  und  $\hat{\sigma}^2$  sind Funktionen von  $\vec{Y}$ . Da  $Y$  und  $\vec{Y}$  unabhängig sind, sind auch  $Y$  und  $((\hat{\beta}_0, \hat{\beta}_1), \hat{\sigma}^2)$  unabhängig und somit auch  $Y$ ,  $(\hat{\beta}_0, \hat{\beta}_1)^T$  und  $\hat{\sigma}^2$ .
3. Aus den Gleichungen (4.1.28) und (4.1.15) folgt

$$\begin{aligned}
 \text{Var}(Y - \hat{y}(x)) &= \text{Var}(Y) + \text{Var}(\hat{y}(x)) = \sigma^2 + \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}} \right) \\
 &= \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}} \right) \\
 Y - \hat{y}(x) &= Y - (\hat{\beta}_0 + \hat{\beta}_1 x) \sim \mathfrak{N} \left( 0, \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}} \right) \right)
 \end{aligned} \tag{4.1.32}$$

4. Somit ergibt sich

$$\frac{Y - \hat{y}(x)}{\hat{\sigma} \sqrt{\left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}} \right)}} \sim t_{n-2} \tag{4.1.33}$$

Hieraus folgt, dass

$$V(\vec{Y}) = \left[ \hat{y}(x) - t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma} h(x), \hat{y}(x) + t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma} h(x) \right] \tag{4.1.34}$$

wobei

$$h(x) = \sqrt{\left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}} \right)} \tag{4.1.35}$$

ein  $(1 - \alpha)$ -Vorhersageintervall für  $Y = \beta_0 + \beta_1 x + \epsilon$  darstellt.

## 4.2 Simultane Konfidenzintervalle und gemeinsame Konfidenzbereiche

### Definition 4.2.1. Simultane Konfidenzintervalle für eindimensionale schätzbare Funktionen

Es sei  $L$  eine Menge eindimensionaler schätzbarer Funktionen des Parametervektors  $\vec{\beta}$  des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells und  $(K_\psi)_{\psi \in L}$  eine Familie von Intervallschätzfunktionen. Das heißt,  $K_\psi$  ist eine Abbildung auf dem Stichprobenraum  $(\mathbb{R}^n)$ , deren Bilder Intervalle sind. Dann ist  $(K_\psi)_{\psi \in L}$  eine Familie von simultanen Konfidenzintervallen zum Niveau  $(1 - \alpha)$  ( $0 < \alpha < 1$ ) für die  $\psi \in L$ , wenn

$$P_{\vec{\beta}, \sigma^2}(\psi(\vec{\beta}) \in K_\psi(\vec{Y}) \mid \forall \psi \in L) \geq 1 - \alpha \quad \forall \vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0. \quad (4.2.1)$$

### Bemerkung 1.

Ist  $L = \{\psi_1, \dots, \psi_m\}$  eine endliche Menge eindimensionaler schätzbarer Funktionen des Parametervektors  $\vec{\beta}$  des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells und  $(K_{\psi_i})_{i=1, \dots, m}$  eine Familie von simultanen Konfidenzintervallen zum Niveau  $(1 - \alpha)$  ( $0 < \alpha < 1$ ) für  $\psi_1, \dots, \psi_m$ , dann ist

$$K(\vec{Y}) = \prod_{i=1}^m K_{\psi_i}(\vec{Y}) = K_{\psi_1}(\vec{Y}) \times \dots \times K_{\psi_m}(\vec{Y}) \quad (4.2.2)$$

ein gemeinsamer Konfidenzbereich für  $\psi_1, \dots, \psi_m$  zum Niveau  $(1 - \alpha)$  ( $0 < \alpha < 1$ ).

### Satz 4.2.1. Simultane Konfidenzintervalle - Bonferroni-Methode

Es seien  $\psi_1(\vec{\beta}) = \vec{c}_1^T \vec{\beta}, \dots, \psi_m(\vec{\beta}) = \vec{c}_m^T \vec{\beta}$  eindimensionale schätzbare Funktionen des Parametervektors  $\vec{\beta}$  des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells mit den BLE-Schätzern  $\hat{\psi}_1 = \vec{c}_1^T \hat{\beta} = \vec{a}_1^T \vec{Y}, \dots, \hat{\psi}_m = \vec{c}_m^T \hat{\beta} = \vec{a}_m^T \vec{Y}$  und der Schätzung der Varianz  $\hat{\sigma}^2 = \frac{1}{n-r}(\vec{Y} - X\hat{\beta})^T(\vec{Y} - X\hat{\beta})$ . Hierbei ist  $\hat{\beta}$  eine beliebige Lösung der Normalgleichungen und die Vektoren  $\vec{a}_i$  liegen im Spaltenraum von  $X$ . Dann sind die Intervalle

$$K_{\psi_i}(\vec{Y}) = [\hat{\psi}_i - t_{n-r; 1-\frac{\alpha}{2m}} \hat{\sigma}_{\hat{\psi}_i}, \hat{\psi}_i + t_{n-r; 1-\frac{\alpha}{2m}} \hat{\sigma}_{\hat{\psi}_i}]$$

simultane Konfidenzintervalle für  $\psi_1, \dots, \psi_m$  zum Niveau  $1 - \alpha$  ( $0 < \alpha < 1$ ). Hierbei ist

$$\hat{\sigma}_{\hat{\psi}_i}^2 = \hat{\sigma}^2 \vec{c}_i^T (X^T X)^{-1} \vec{c}_i = \hat{\sigma}^2 \vec{a}_i^T \vec{a}_i \quad (4.2.3)$$

die erwartungstreue Schätzung für  $\sigma_{\psi}^2$ .

**Beweis:**

Nach Gleichung (4.1.4) ist  $K_{\psi_i}$  jeweils ein  $(1 - \frac{\alpha}{m})$ -Konfidenzintervall für  $\psi_i$  ( $i = 1, \dots, m$ ). Hieraus folgt

$$\begin{aligned}
 P_{\vec{\beta}, \sigma^2}(\psi_i(\vec{\beta}) \in K_{\psi_i}(\vec{Y}), i = 1, \dots, m) \\
 &= P_{\vec{\beta}, \sigma^2} \left( \bigcap_{i=1}^m \{\psi_i(\vec{\beta}) \in K_{\psi_i}(\vec{Y})\} \right) = 1 - P_{\vec{\beta}, \sigma^2} \left( \bigcup_{i=1}^m \overline{\{\psi_i(\vec{\beta}) \in K_{\psi_i}(\vec{Y})\}} \right) \\
 &\geq 1 - \sum_{i=1}^m P_{\vec{\beta}, \sigma^2} \left( \overline{\{\psi_i(\vec{\beta}) \in K_{\psi_i}(\vec{Y})\}} \right) = 1 - \sum_{i=1}^m \left( 1 - P_{\vec{\beta}, \sigma^2} \left( \{\psi_i(\vec{\beta}) \in K_{\psi_i}(\vec{Y})\} \right) \right) \\
 &\geq 1 - \sum_{i=1}^m \left( 1 - \left( 1 - \frac{\alpha}{m} \right) \right) = 1 - \alpha
 \end{aligned} \tag{4.2.4}$$

**Satz 4.2.2. Konfidenzellipse für eine  $q$ -dimensionale schätzbare Funktion**

Es sei  $\hat{\psi} = C\hat{\beta} = A\vec{Y}$  der BLE-Schätzer für die  $q$ -dimensionale schätzbare Funktion  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  des Parametervektors  $\vec{\beta}$  des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells mit  $rg(C) = q$  und  $\hat{\sigma}^2 = \frac{1}{n-r}(\vec{Y} - X\hat{\beta})^T(\vec{Y} - X\hat{\beta})$  die erwartungstreue Schätzung der Varianz  $\sigma^2$ . Dann ist  $q \leq r = rg(X)$  und  $B = AA^T = C(X^T X)^{-1}C^T$  und damit auch  $\Sigma_{\hat{\psi}} = \sigma^2 B$  regulär, und für einen beliebigen Vektor  $(\vec{\beta}^T, \sigma^2)^T \in \Omega$  gelten die folgenden Aussagen:

1.  $\frac{1}{\sigma^2}(\hat{\psi} - \vec{\psi}(\vec{\beta}))^T B^{-1}(\hat{\psi} - \vec{\psi}(\vec{\beta})) \sim \chi_q^2$
2.  $\frac{n-r}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} \hat{\epsilon}^T \hat{\epsilon} = \frac{1}{\sigma^2} (\vec{Y} - X\hat{\beta})^T (\vec{Y} - X\hat{\beta}) \sim \chi_{n-r}^2$
3.  $(\hat{\psi} - \vec{\psi}(\vec{\beta}))^T B^{-1}(\hat{\psi} - \vec{\psi}(\vec{\beta}))$  und  $\hat{\sigma}^2$  sind unabhängig.
4.  $\frac{(\hat{\psi} - \vec{\psi}(\vec{\beta}))^T B^{-1}(\hat{\psi} - \vec{\psi}(\vec{\beta}))}{q\hat{\sigma}^2} = \frac{\frac{1}{q\sigma^2}(\hat{\psi} - \vec{\psi}(\vec{\beta}))^T B^{-1}(\hat{\psi} - \vec{\psi}(\vec{\beta}))}{\frac{1}{(n-r)\sigma^2}(\vec{Y} - X\hat{\beta})^T (\vec{Y} - X\hat{\beta})} \sim F_{q, n-r}$
5.  $P_{\vec{\beta}, \sigma^2} \left( \frac{(\hat{\psi} - \vec{\psi}(\vec{\beta}))^T B^{-1}(\hat{\psi} - \vec{\psi}(\vec{\beta}))}{q\hat{\sigma}^2} \leq F_{q, n-r; 1-\alpha} \right) = 1 - \alpha \quad \forall \vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0$

6. Das zufällige  $q$ -dimensionale Hyperellipsoid

$$\begin{aligned}
 K(\vec{Y}) &= \left\{ \vec{\psi} \in \mathbb{R}^q \mid \frac{(\hat{\psi} - \vec{\psi})^T B^{-1}(\hat{\psi} - \vec{\psi})}{q\hat{\sigma}^2} \leq F_{q, n-r; 1-\alpha} \right\} \\
 &= \left\{ \vec{\psi} \in \mathbb{R}^q \mid (\hat{\psi} - \vec{\psi})^T B^{-1}(\hat{\psi} - \vec{\psi}) \leq q\hat{\sigma}^2 F_{q, n-r; 1-\alpha} \right\}
 \end{aligned} \tag{4.2.5}$$

ist ein  $(1 - \alpha)$ -Konfidenzbereich für  $\vec{\psi}$  und wird als  $(1 - \alpha)$ -Konfidenzellipse für  $\vec{\psi}$  bezeichnet.

**Beweis:**

Da  $A\vec{Y}$  eine erwartungstreue Schätzung für  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  ist, gilt nach Korollar 2.2 zu Satz 2.1.4  $C = AX$ . Hieraus folgt  $q = rg(C) \leq rg(X) = r$  sowie, da  $B$  eine  $(q \times q)$ -Matrix ist,  $q = rg(C) \leq rg(A) = rg(AA^T) = rg(B) \leq q$ . Somit ist  $B$  regulär. Nach Satz 3.3.2 gilt

$$\hat{\psi} \sim \mathfrak{N}(\vec{\psi}(\vec{\beta}), \sigma^2 B) \quad (4.2.6)$$

Hieraus folgt mit Satz 3.2.3 die Aussage 1.

Aus Satz 3.3.2 folgt direkt die Aussage 2 und, da  $(\hat{\psi} - \vec{\psi}(\vec{\beta}))^T B^{-1}(\hat{\psi} - \vec{\psi}(\vec{\beta}))$  eine Funktion von  $\hat{\psi}$  ist und  $\hat{\psi}$  und  $\hat{\sigma}^2$  unabhängig sind, die Aussage 3.

Die weiteren Aussagen ergeben sich direkt aus der Definition der  $F$ -Verteilung.

**Beispiel 4.2.1. Gemeinsames Konfidenzellipsoid für die Parameter bei vollem Spaltenrang**

Die Designmatrix des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells besitze vollen Spaltenrang. Dann ist der Parametervektor  $\vec{\beta}$  eine  $k$ -dimensionale schätzbare Funktion. Da

$$\hat{\beta} = (X^T X)^{-1} X^T \vec{Y} \sim \mathfrak{N}(\vec{\beta}, \sigma^2 (X^T X)^{-1}) \quad (4.2.7)$$

ergibt sich mit  $\vec{\psi}(\vec{\beta}) = \vec{\beta}$ ,  $B = (X^T X)^{-1}$  und  $q = k$  aus Gleichung (4.2.5) der folgende  $(1 - \alpha)$ -Konfidenzbereich für  $\vec{\beta}$ :

$$K(\vec{Y}) = \left\{ \vec{\beta} \in \mathbb{R}^k \mid (\hat{\beta} - \vec{\beta})^T X^T X (\hat{\beta} - \vec{\beta}) \leq k \hat{\sigma}^2 F_{k, n-k; 1-\alpha} \right\} \quad (4.2.8)$$

**Satz 4.2.3. Simultane Konfidenzintervalle - Scheffé-Methode**

Es sei  $\mathcal{L}_q$  ein  $q$ -dimensionaler linearer Teilraum des Raumes der eindimensionalen schätzbaren Funktionen des Parametervektors  $\vec{\beta}$  des durch Gleichung (4.1.1) gegebenen klassischen linearen Modells. Dann ist  $(K_\psi)_{\psi \in \mathcal{L}_q}$  mit

$$\begin{aligned} K_\psi(\vec{Y}) &= \left\{ \psi \in \mathbb{R} \mid \hat{\psi} - \sqrt{q F_{q, n-r; 1-\alpha}} \hat{\sigma}_{\hat{\psi}} \leq \psi \leq \hat{\psi} + \sqrt{q F_{q, n-r; 1-\alpha}} \hat{\sigma}_{\hat{\psi}} \right\} \\ &= [\hat{\psi} - \sqrt{q F_{q, n-r; 1-\alpha}} \hat{\sigma}_{\hat{\psi}}, \hat{\psi} + \sqrt{q F_{q, n-r; 1-\alpha}} \hat{\sigma}_{\hat{\psi}}] \end{aligned} \quad (4.2.9)$$

eine Familie von simultanen Konfidenzintervallen zum Niveau  $(1 - \alpha)$  ( $0 < \alpha < 1$ ) für die  $\psi \in \mathcal{L}_q$ . Hierbei ist  $\hat{\psi} = \vec{c}^T \hat{\beta} = \vec{a}^T \vec{Y}$  der BLE-Schätzer für  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  mit einer beliebigen Lösung  $\hat{\beta}$  der Normalgleichungen und einem im Spaltenraum von  $X$  liegenden Vektor  $\vec{a}$  und

$$\hat{\sigma}_{\hat{\psi}}^2 = \hat{\sigma}^2 \vec{c}^T (X^T X)^{-1} \vec{c} = \hat{\sigma}^2 \vec{a}^T \vec{a} \quad (4.2.10)$$

die erwartungstreue Schätzung der Varianz von  $\hat{\psi}$ .

**Beweis:**

Es seien  $\phi_1, \dots, \phi_q$  linear unabhängige Funktionen, die den Raum  $\mathcal{L}_q$  aufspannen, und  $\hat{\phi} = C\hat{\beta} = A\vec{Y}$  der BLE-Schätzer für die  $q$ -dimensionale schätzbare Funktion  $\vec{\phi}(\vec{\beta}) = (\phi_1(\vec{\beta}), \dots, \phi_q(\vec{\beta}))^T =$



$C\vec{\beta}$  mit  $rg(C) = q$  und  $\Sigma_{\hat{\phi}} = \sigma^2 B$ . Dann ist  $B = AA^T = C(X^T X)^{-1}C^T$  regulär (Satz 4.2.2), und es gilt nach Aussage 5 in Satz 4.2.2

$$P_{\vec{\beta}, \sigma^2} \left( \frac{1}{\hat{\sigma}^2} (\hat{\phi} - \vec{\phi}(\vec{\beta}))^T B^{-1} (\hat{\phi} - \vec{\phi}(\vec{\beta})) \leq qF_{q, n-r; 1-\alpha} \right) = 1 - \alpha \quad \forall \vec{\beta} \in \mathbb{R}^k, \sigma^2 > 0. \quad (4.2.11)$$

Eine schätzbare Funktion  $\psi \in \mathcal{L}_q \setminus \{\mathbf{o}\}$  besitzt die Darstellung  $\psi = \vec{d}^T \vec{\phi}$  mit  $\vec{d} \in \mathbb{R}^q \setminus \{\mathbf{o}\}$ . Für den BLE-Schätzer  $\hat{\psi} = \vec{d}^T \hat{\phi}$  von  $\psi = \vec{d}^T \vec{\phi}$  ergibt sich

$$\sigma_{\hat{\psi}}^2 = \vec{d}^T \Sigma_{\hat{\phi}} \vec{d} = \sigma^2 \vec{d}^T B \vec{d} \quad (4.2.12)$$

und somit

$$\hat{\sigma}_{\hat{\psi}}^2 = \hat{\sigma}^2 \vec{d}^T B \vec{d}. \quad (4.2.13)$$

Hieraus folgt

$$\frac{(\hat{\psi} - \psi(\vec{\beta}))^2}{\hat{\sigma}_{\hat{\psi}}^2} = \frac{(\vec{d}^T (\hat{\phi} - \vec{\phi}(\vec{\beta})))^2}{\hat{\sigma}^2 \vec{d}^T B \vec{d}} = \frac{(\vec{d}^T B B^{-1} (\hat{\phi} - \vec{\phi}(\vec{\beta})))^2}{\hat{\sigma}^2 \vec{d}^T B \vec{d}}. \quad (4.2.14)$$

Da  $B$  positiv definit und symmetrisch ist, wird durch

$$\langle \vec{x}, \vec{y} \rangle_B := \vec{x}^T B \vec{y}, \quad \vec{x}, \vec{y} \in \mathbb{R}^n \quad (4.2.15)$$

ein Skalarprodukt im  $\mathbb{R}^n$  definiert. Mit der Ungleichung von Cauchy-Schwarz (Korollar 2 zu Satz 1.2.3) ergibt sich dann mit  $\vec{y} := B^{-1}(\hat{\phi} - \vec{\phi}(\vec{\beta}))$  und  $\|\vec{x}\|_B := \langle \vec{x}, \vec{x} \rangle_B^{1/2}$

$$\begin{aligned} \frac{(\hat{\psi} - \psi(\vec{\beta}))^2}{\hat{\sigma}_{\hat{\psi}}^2} &= \frac{\langle \vec{d}, \vec{y} \rangle_B^2}{\hat{\sigma}^2 \langle \vec{d}, \vec{d} \rangle_B} \leq \frac{\|\vec{d}\|_B^2 \cdot \|\vec{y}\|_B^2}{\hat{\sigma}^2 \|\vec{d}\|_B^2} = \frac{\|\vec{y}\|_B^2}{\hat{\sigma}^2} \\ „ = “ \quad &\Leftrightarrow \quad \vec{d} = c\vec{y} \quad \Leftrightarrow \quad \vec{d} = cB^{-1}(\hat{\phi} - \vec{\phi}(\vec{\beta})) \\ &\Leftrightarrow \quad \psi = c(\hat{\phi} - \vec{\phi}(\vec{\beta}))^T B^{-1} \vec{\phi}, \end{aligned} \quad (4.2.16)$$

hieraus folgt

$$\begin{aligned} \max_{\psi \in \mathcal{L}_q \setminus \{\mathbf{o}\}} \frac{(\hat{\psi} - \psi(\vec{\beta}))^2}{\hat{\sigma}_{\hat{\psi}}^2} &= \frac{\vec{y}^T B \vec{y}}{\hat{\sigma}^2} = \frac{1}{\hat{\sigma}^2} (\hat{\phi} - \vec{\phi}(\vec{\beta}))^T (B^{-1})^T B B^{-1} (\hat{\phi} - \vec{\phi}(\vec{\beta})) \\ &= \frac{1}{\hat{\sigma}^2} (\hat{\phi} - \vec{\phi}(\vec{\beta}))^T B^{-1} (\hat{\phi} - \vec{\phi}(\vec{\beta})). \end{aligned} \quad (4.2.17)$$

Somit erhalten wir mit Gleichung (4.2.11)

$$\begin{aligned} 1 - \alpha &= P_{\vec{\beta}, \sigma^2} \left( \frac{1}{\hat{\sigma}^2} (\hat{\phi} - \vec{\phi}(\vec{\beta}))^T B^{-1} (\hat{\phi} - \vec{\phi}(\vec{\beta})) \leq qF_{q, n-r; 1-\alpha} \right) \\ &= P_{\vec{\beta}, \sigma^2} \left( \max_{\psi \in \mathcal{L}_q \setminus \{\mathbf{o}\}} \frac{(\hat{\psi} - \psi(\vec{\beta}))^2}{\hat{\sigma}_{\hat{\psi}}^2} \leq qF_{q, n-r; 1-\alpha} \right) \\ &= P_{\vec{\beta}, \sigma^2} \left( \hat{\psi} - \sqrt{qF_{q, n-r; 1-\alpha}} \hat{\sigma}_{\hat{\psi}} \leq \psi(\vec{\beta}) \leq \hat{\psi} + \sqrt{qF_{q, n-r; 1-\alpha}} \hat{\sigma}_{\hat{\psi}} \quad \forall \psi \in \mathcal{L}_q \right) \end{aligned} \quad (4.2.18)$$

**Beispiel 4.2.2. Konfidenzband für die Regressionsgerade**

Für das durch die Gleichungen

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (i = 1, \dots, n), \quad \sigma^2 > 0 \quad (4.2.19)$$

$$\epsilon_1, \dots, \epsilon_n \quad \text{stochastisch unabhängig}$$

gegebene einfache lineare Regressionsmodell ergeben sich mit den Ergebnissen aus Beispiel 4.1.5 die folgenden Aussagen.

Die Bonferroni-Methode ergibt für die Werte der Regressionsgeraden  $y(x) = \beta_0 + \beta_1 x$  an den Stellen  $\tilde{x}_1, \dots, \tilde{x}_m$  die simultanen  $(1 - \alpha)$ -Konfidenzintervalle

$$K_{y(\tilde{x}_j)}(\vec{Y}) = \left[ \hat{y}(\tilde{x}_j) - t_{n-2; 1-\frac{\alpha}{2m}} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(\tilde{x}_j - \bar{x})^2}{ns_{xx}}}, \hat{y}(\tilde{x}_j) + t_{n-2; 1-\frac{\alpha}{2m}} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(\tilde{x}_j - \bar{x})^2}{ns_{xx}}} \right] \quad (4.2.20)$$

Die Scheffé-Methode ergibt, da  $L = \{\psi \mid \psi((\beta_0, \beta_1)^T) = \beta_0 + \beta_1 x, x \in \mathbb{R}\} \subseteq \mathcal{L}_2 = \{\psi \mid \psi((\beta_0, \beta_1)^T) = c_1 \beta_0 + c_2 \beta_1, (c_1, c_2)^T \in \mathbb{R}^2\}$ , für die Werte der Regressionsgeraden  $y(x)$  ( $x \in \mathbb{R}$ ) die folgende Familie simultaner Konfidenzintervalle zum Niveau  $(1 - \alpha)$

$$K_{y(x)}(\vec{Y}) = \left[ \hat{y}(x) - \hat{\sigma}_{\hat{y}(x)} \sqrt{2F_{2, n-2; 1-\alpha}}, \hat{y}(x) + \hat{\sigma}_{\hat{y}(x)} \sqrt{2F_{2, n-2; 1-\alpha}} \right] \quad (4.2.21)$$

mit

$$\hat{\sigma}_{\hat{y}(x)} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{ns_{xx}}} \quad (4.2.22)$$

Die Menge

$$B_{1-\alpha}(\vec{Y}) = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid \hat{y}(x) - \hat{\sigma}_{\hat{y}(x)} \sqrt{2F_{2, n-2; 1-\alpha}} \leq y \leq \hat{y}(x) + \hat{\sigma}_{\hat{y}(x)} \sqrt{2F_{2, n-2; 1-\alpha}} \right\} \quad (4.2.23)$$

heißt Konfidenzband zum Niveau  $1 - \alpha$  für die Regressionsgerade  $y(x) = \beta_0 + \beta_1 x$ .

Die Menge

$$B_{1-\alpha}^*(\vec{Y}) = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid \hat{y}(x) - \hat{\sigma}_{\hat{y}(x)} t_{n-2; 1-\frac{\alpha}{2}} \leq y \leq \hat{y}(x) + \hat{\sigma}_{\hat{y}(x)} t_{n-2; 1-\frac{\alpha}{2}} \right\} \quad (4.2.24)$$

heißt Konfidenzband zum Niveau  $1 - \alpha$  für einen einzelnen Wert der Regressionsgeraden  $y(x) = \beta_0 + \beta_1 x$ .

**Beispiel 4.2.3. Konfidenzband für die Regressionsfunktion bei vollem Spaltenrang**

Wir betrachten im Folgenden das klassische lineare Modell mit Achsenabschnitt für den  $n$ -dimensionalen Beobachtungsvektor  $\vec{Y} = (Y_1, \dots, Y_n)$

$$\begin{aligned} Y_i &= \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k + \epsilon_i, \quad i = 1, \dots, n \\ \vec{Y} &= X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = k + 1 < n, \\ \vec{\beta} &= (\beta_0, \beta_1, \dots, \beta_k)^T \in \mathbb{R}^{k+1}, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^{k+1} \times \mathbb{R}_{>0} \end{aligned} \quad (4.2.25)$$

Die Scheffé-Methode ergibt, da  $L = \{\psi \mid \psi(\vec{\beta}) = \beta_0 + x_1\beta_1 + \dots + x_k\beta_k, \vec{x} = (x_1, \dots, x_k)^T \in \mathbb{R}^k\} \subseteq \mathcal{L}_{k+1} = \{\psi \mid \psi(\vec{\beta}) = \vec{c}^T \vec{\beta}, \vec{c} \in \mathbb{R}^{k+1}\}$ , für die Werte der Regressionsfunktion  $y(\vec{x}) = \beta_0 + x_1\beta_1 + \dots + x_k\beta_k$  ( $\vec{x} = (x_1, \dots, x_k)^T \in \mathbb{R}^k$ ) die folgende Familie simultaner Konfidenzintervalle zum Niveau  $(1 - \alpha)$

$$K_{y(\vec{x})}(\vec{Y}) = \left[ \hat{y}(\vec{x}) - \hat{\sigma}_{\hat{y}(\vec{x})} \sqrt{(k+1)F_{k+1, n-(k+1); 1-\alpha}}, \hat{y}(\vec{x}) + \hat{\sigma}_{\hat{y}(\vec{x})} \sqrt{(k+1)F_{k+1, n-(k+1); 1-\alpha}} \right] \quad (4.2.26)$$

mit

$$\hat{\sigma}_{\hat{y}(\vec{x})} = \hat{\sigma} \sqrt{(1, \vec{x}^T)(X^T X)^{-1}(1, \vec{x}^T)^T} \quad (4.2.27)$$



# Kapitel 5

## Hypothesentests

### 5.1 F-Test für eine testbare Hypothese

#### Definition 5.1.1. Testbare Hypothese

Eine Hypothese über die Parameter des klassischen linearen Modells für den  $n$ -dimensionalen Beobachtungsvektors  $\vec{Y}$

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^k \times \mathbb{R}_{>0} \quad (5.1.1)$$

der Form

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{c} \quad (5.1.2)$$

heißt testbar genau dann, wenn  $\vec{\psi}$  schätzbar ist und  $\vec{c}$  im Spaltenraum der Matrix  $C$  liegt.

Der durch  $H$  festgelegte Unterraum des Parameterraumes  $\Omega$  wird mit

$$\omega = \omega_H = \left\{ \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega \mid \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{c} \right\} \quad (5.1.3)$$

bezeichnet.

#### Bemerkung 1.

Das (durch die folgenden Gleichungen gegebene) Modell

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \omega \quad (5.1.4)$$

das vorliegt, wenn die durch Gleichung (5.1.2) gegebene testbare Hypothese  $H$  für das durch Gleichung (5.1.1) gegebene lineare Modell erfüllt ist, wird als (durch  $H$  gegebenes) reduziertes Modell und das Modell selbst als vollständiges (lineares) Modell bezeichnet. Für  $\vec{c} = \vec{0}$  ist das reduzierte Modell ein lineares Modell.

**Bemerkung 2.**

Eine testbare Hypothese der Form

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{c} \quad (5.1.5)$$

für das durch Gleichung (5.1.1) gegebene klassische lineare Modell kann durch die Transformationen

$$\vec{\beta}^* = \vec{\beta} - \tilde{\beta}, \quad \vec{Y}^* = \vec{Y} - X\tilde{\beta} \quad (5.1.6)$$

wobei  $\tilde{\beta}$  eine beliebige Lösung von  $C\vec{\beta} = \vec{c}$  ist, auf eine testbare Hypothese der Form

$$H^* : \quad \vec{\psi}(\vec{\beta}^*) = C\vec{\beta}^* = \vec{0} \quad (5.1.7)$$

für das lineare Modell

$$\vec{Y}^* = X\vec{\beta}^* + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \quad \begin{pmatrix} \vec{\beta}^* \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^k \times \mathbb{R}_{>0} \quad (5.1.8)$$

zurückgeführt werden. Die Transformationen entsprechen einer Verschiebung des Ursprungs des Koordinatensystems des Parameterraumes nach  $\tilde{\beta}$  und des Beobachtungsraumes nach  $X\tilde{\beta}$ .

Dann ist  $\hat{\beta}_\Omega$  eine Lösung der Normalgleichungen des durch Gleichung (5.1.1) gegebene linearen Modells genau dann, wenn  $\hat{\beta}_\Omega^* = \hat{\beta}_\Omega - \tilde{\beta}$  eine Lösung der Normalgleichungen des durch Gleichung (5.1.8) gegebenen linearen Modells ist.

$\hat{\beta}_{\omega_H}$  ist eine Lösung der Minimierungsaufgabe

$$(\vec{Y} - X\hat{\beta}_{\omega_H})^T (\vec{Y} - X\hat{\beta}_{\omega_H}) = \min_{(\vec{\beta}^T, \sigma^2)^T \in \omega_H} (\vec{Y} - X\vec{\beta})^T (\vec{Y} - X\vec{\beta}) \quad (5.1.9)$$

genau dann, wenn  $\hat{\beta}_{\omega_{H^*}}^* = \hat{\beta}_{\omega_H} - \tilde{\beta}$  eine Lösung der Minimierungsaufgabe

$$(\vec{Y}^* - X\hat{\beta}_{\omega_{H^*}}^*)^T (\vec{Y}^* - X\hat{\beta}_{\omega_{H^*}}^*) = \min_{((\vec{\beta}^*)^T, \sigma^2)^T \in \omega_{H^*}} (\vec{Y}^* - X\vec{\beta}^*)^T (\vec{Y}^* - X\vec{\beta}^*) \quad (5.1.10)$$

ist.

Die Residuenquadratsummen des ursprünglichen Modells

$$S_\Omega := (\vec{Y} - X\hat{\beta}_\Omega)^T (\vec{Y} - X\hat{\beta}_\Omega), \quad S_{\omega_H} := (\vec{Y} - X\hat{\beta}_{\omega_H})^T (\vec{Y} - X\hat{\beta}_{\omega_H}) \quad (5.1.11)$$

und des transformierten Modells

$$S_\Omega^* := (\vec{Y}^* - X\hat{\beta}_\Omega^*)^T (\vec{Y}^* - X\hat{\beta}_\Omega^*), \quad S_{\omega_{H^*}}^* := (\vec{Y}^* - X\hat{\beta}_{\omega_{H^*}}^*)^T (\vec{Y}^* - X\hat{\beta}_{\omega_{H^*}}^*) \quad (5.1.12)$$

stimmen überein

$$S_\Omega = S_\Omega^*, \quad S_{\omega_H} = S_{\omega_{H^*}}^* \quad (5.1.13)$$

**Definition 5.1.2. Likelihood-Quotiententest**

Es sei

$$L(\vec{\beta}, \sigma^2; \vec{y}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2}(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta})} \quad (5.1.14)$$

die Likelihoodfunktion für das durch Gleichung (5.1.1) gegebene klassische lineare Modell, dann ist

$$\lambda(\vec{y}) = \frac{\max_{(\vec{\beta}^T, \sigma^2)^T \in \omega} L(\vec{\beta}, \sigma^2; \vec{y})}{\max_{(\vec{\beta}^T, \sigma^2)^T \in \Omega} L(\vec{\beta}, \sigma^2; \vec{y})} \quad (5.1.15)$$

die Testgröße des Likelihood-Quotiententests für eine durch Gleichung (5.1.2) gegebene testbare Hypothese  $H$ . Die Hypothese wird zum Niveau  $\alpha$  abgelehnt genau dann, wenn

$$\lambda(\vec{y}) < c_\alpha \quad (5.1.16)$$

wobei  $c_\alpha$  das  $\alpha$ -Quantil der Verteilung von  $\lambda(\vec{Y})$  bei Vorliegen von  $H$  ist.

**Satz 5.1.1. Likelihood-Quotiententest**

Es sei  $S_\Omega$  die Residuenquadratsumme des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells und  $S_\omega$  die Residuenquadratsumme des durch eine testbare Hypothese  $H$ , definiert in Gleichung (5.1.2), mit  $\text{rg}(C) = q$  gegebenen reduzierten Modells (siehe Gleichung (5.1.11)). Dann ergibt sich die Testgröße des Likelihoodquotiententests für  $H$  zu

$$\begin{aligned} \lambda(\vec{y}) &= \left( \frac{S_\omega}{S_\Omega} \right)^{-\frac{n}{2}} = \left( 1 + \frac{q}{n-r} \frac{\frac{1}{q}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} \right)^{-\frac{n}{2}} \\ &= \left( 1 + \frac{q}{n-r} F \right)^{-\frac{n}{2}} \end{aligned} \quad (5.1.17)$$

mit

$$F = \frac{\frac{1}{q}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} \quad (5.1.18)$$

Somit führt der Likelihood-Quotiententest zum Niveau  $\alpha$  zur Ablehnung der Hypothese  $H$  genau dann, wenn  $F > f_{1-\alpha}$ , wobei  $f_{1-\alpha}$  das  $(1 - \alpha)$ -Quantil der Verteilung von  $F$  bei Vorliegen von  $H$  ist.

**Beweis:**

Es seien  $\hat{\beta}_\Omega$ ,  $\hat{\sigma}_{\Omega,ML}^2$  und  $\hat{Y}_\Omega = \hat{\mu}_\Omega = X\hat{\beta}_\Omega$  Maximum-Likelihood-Schätzungen für  $\vec{\beta}$ ,  $\sigma^2$  und  $E[\vec{Y}] = \vec{\mu}_{\vec{Y}} = X\vec{\beta}$  und  $S_\Omega = (\vec{Y} - \hat{Y}_\Omega)^T(\vec{Y} - \hat{Y}_\Omega)$  die Residuenquadratsumme des vollständigen linearen Modells. Dann ergibt sich (s. Beweis zu Satz 3.3.1)

$$\max_{(\vec{\beta}^T, \sigma^2)^T \in \Omega} L(\vec{\beta}, \sigma^2; \vec{Y}) = \frac{1}{(\hat{\sigma}_{\Omega,ML}\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\hat{\sigma}_{\Omega,ML}^2}(\vec{Y}-X\hat{\beta}_\Omega)^T(\vec{Y}-X\hat{\beta}_\Omega)} = \frac{1}{(\hat{\sigma}_{\Omega,ML}\sqrt{2\pi})^n} \cdot e^{-\frac{n}{2}} \quad (5.1.19)$$

Für das durch  $H$  reduzierte Modell ergibt sich

$$\max_{(\vec{\beta}^T, \sigma^2)^T \in \omega} L(\vec{\beta}, \sigma^2; \vec{Y}) = \max_{(\vec{\beta}^T, \sigma^2)^T \in \omega} \frac{1}{(\sigma \sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2}(\vec{Y} - X\vec{\beta})^T(\vec{Y} - X\vec{\beta})} \quad (5.1.20)$$

Die Likelihoodfunktion wird maximiert bezüglich  $\vec{\beta}$  genau dann, wenn

$$(\vec{Y} - X\vec{\beta})^T(\vec{Y} - X\vec{\beta}) \quad (5.1.21)$$

bezüglich  $\vec{\beta}$  minimiert wird. Daher sind die Lösungen  $\hat{\beta}_\omega$  der Minimierungsaufgabe

$$(\vec{Y} - X\hat{\beta}_\omega)^T(\vec{Y} - X\hat{\beta}_\omega) = \min_{(\vec{\beta}^T, \sigma^2)^T \in \omega} (\vec{Y} - X\vec{\beta})^T(\vec{Y} - X\vec{\beta}) \quad (5.1.22)$$

genau die Maximum-Likelihood-Schätzwerte von  $\vec{\beta}$  für das reduzierte Modell. Die Maximum-Likelihood-Schätzung  $\hat{\sigma}_{\omega, ML}^2$  von  $\sigma^2$  für das reduzierte Modell ergibt sich analog zu Gleichung (3.3.15) zu

$$\hat{\sigma}_{\omega, ML}^2 = \frac{1}{n}(\vec{Y} - X\hat{\beta}_\omega)^T(\vec{Y} - X\hat{\beta}_\omega) \quad (5.1.23)$$

Des weiteren sei  $\hat{Y}_\omega = \hat{\mu}_\omega = X\hat{\beta}_\omega$  die Maximum-Likelihood-Schätzung von  $E[\vec{Y}] = \mu_{\vec{Y}} = X\vec{\beta}$  und  $S_\omega = (\vec{Y} - \hat{Y}_\omega)^T(\vec{Y} - \hat{Y}_\omega)$  die Residuenquadratsumme des reduzierten linearen Modells. Dann ergibt sich

$$\max_{(\vec{\beta}^T, \sigma^2)^T \in \omega} L(\vec{\beta}, \sigma^2; \vec{Y}) = \frac{1}{(\hat{\sigma}_{\omega, ML} \sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\hat{\sigma}_{\omega, ML}^2}(\vec{Y} - X\hat{\beta}_\omega)^T(\vec{Y} - X\hat{\beta}_\omega)} = \frac{1}{(\hat{\sigma}_{\omega, ML} \sqrt{2\pi})^n} \cdot e^{-\frac{n}{2}} \quad (5.1.24)$$

Die Testgröße des Likelihood-Quotienten-Tests lautet somit

$$\lambda(\vec{Y}) = \left( \frac{\hat{\sigma}_{\omega, ML}}{\hat{\sigma}_{\Omega, ML}} \right)^{-n} = \left( \frac{S_\omega}{S_\Omega} \right)^{-\frac{n}{2}} \quad (5.1.25)$$

Da die Testgröße des Likelihood-Quotienten-Tests eine streng monoton fallende Funktion von  $F$  ist, ergibt sich die alternative Form des Likelihood-Quotiententests.

### Satz 5.1.2.

Es sei

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{0} \quad (5.1.26)$$

eine testbare Hypothese mit  $rg(C) = q$  des durch Gleichung (5.1.1) gegebenen linearen Modells mit  $rg(X) = r$  ( $r \leq k$ ) und  $\hat{\psi} = A\vec{Y}$  der BLE-Schätzer der  $q$ -dimensionalen schätzbaren Funktion  $\vec{\psi}$ . Hierbei ist  $A = [\vec{a}_1 \ \vec{a}_2 \ \dots \ \vec{a}_q]^T$  eine Matrix, deren Zeilenvektoren  $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_q$  im Spaltenraum von  $X$  liegen (s. Korollar 1 zu Satz 2.1.5), mit  $C = AX$  (s. Korollar 1 zu Satz 2.1.4) und  $rg(A) = rg(C) = q$  (s. Satz 4.2.2).



Der Vektor  $\vec{\mu} = X\vec{\beta}$  ist ein unter der Hypothese möglicher Erwartungswert von  $\vec{Y}$  genau dann, wenn

$$C\vec{\beta} = AX\vec{\beta} = A\vec{\mu} = \vec{0}. \quad (5.1.27)$$

Somit ist der Vektor  $\vec{\mu} \in \mathbb{R}^n$  ein unter der Hypothese möglicher Erwartungswert des linearen Modells genau dann, wenn

$$\vec{\mu} \in \mathfrak{Im}(X) \cap \mathfrak{K}(A) = \mathfrak{Im}(X) \cap \mathfrak{Im}(A^T)^\perp \quad (5.1.28)$$

Es sei

$$\begin{aligned} \mathcal{V}_\Omega &= \mathcal{V}_r = \mathfrak{Im}(X) = \{\vec{\mu} = X\vec{\beta} \mid (\vec{\beta}^T, \sigma^2)^T \in \Omega\} \\ \mathcal{V}_{\Omega-\omega} &= \mathcal{V}_q = \mathfrak{Im}(A^T) = \mathcal{L}(\vec{a}_1, \dots, \vec{a}_q) \subseteq \mathcal{V}_\Omega \\ \mathcal{V}_\omega &= \mathcal{V}_{r-q} = \{\vec{\mu} = X\vec{\beta} \mid (\vec{\beta}^T, \sigma^2)^T \in \omega\} = \{\vec{\mu} \mid \vec{\mu} \in \mathcal{V}_\Omega, \vec{\mu} \in \mathcal{V}_{\Omega-\omega}^\perp\} \end{aligned} \quad (5.1.29)$$

dann gilt

$$\begin{aligned} \mathcal{V}_\Omega &= \mathcal{V}_\omega \oplus \mathcal{V}_{\Omega-\omega}, \quad \mathcal{V}_\omega = \mathcal{V}_\Omega \cap \mathcal{V}_{\Omega-\omega}^\perp \\ \dim(\mathcal{V}_\Omega) &= r, \quad \dim(\mathcal{V}_{\Omega-\omega}) = q, \quad \dim(\mathcal{V}_\omega) = r - q \end{aligned} \quad (5.1.30)$$

Hiermit lautet die parameterfreie Form des vollständigen linearen Modells (s. Bemerkung 2 zu Definition 1.1.1 )

$$\vec{Y} = \vec{\mu} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \vec{\mu} \in \mathcal{V}_\Omega = \mathcal{V}_r, \quad r < n \quad (5.1.31)$$

und die des durch  $H$  reduzierten linearen Modells

$$\vec{Y} = \vec{\mu} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \vec{\mu} \in \mathcal{V}_\omega = \mathcal{V}_{r-q}, \quad r - q < r < n \quad (5.1.32)$$

### Definition 5.1.3. Kanonische Basis

Eine Orthonormalbasis  $\vec{q}_1, \dots, \vec{q}_n$  des  $\mathbb{R}^n$  heißt kanonische Basis für die testbare Hypothese

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{0} \quad (5.1.33)$$

mit  $\text{rg}(C) = q$  des durch Gleichung (5.1.1) gegebenen linearen Modells, dessen Designmatrix  $X$  den Rang  $r$  ( $r \leq k$ ) hat, genau dann, wenn die folgenden Aussagen gelten:

$$\begin{aligned} \mathcal{V}_\Omega &= \mathcal{V}_r = \mathcal{L}(\vec{q}_1, \dots, \vec{q}_r) \\ \mathcal{V}_{\Omega-\omega} &= \mathcal{V}_q = \mathcal{L}(\vec{q}_1, \dots, \vec{q}_q) \\ \mathcal{V}_\omega &= \mathcal{V}_{r-q} = \mathcal{L}(\vec{q}_{q+1}, \dots, \vec{q}_r) \end{aligned} \quad (5.1.34)$$

**Satz 5.1.3.**

Es sei  $\vec{q}_1, \dots, \vec{q}_n$  eine kanonische Basis für die testbare Hypothese

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{0} \quad (5.1.35)$$

mit  $\text{rg}(C) = q$  des durch Gleichung (5.1.1) gegebenen linearen Modells, dessen Designmatrix  $X$  den Rang  $r$  ( $r \leq k$ ) hat. Des weiteren sei

$$\begin{aligned} Q &= \begin{bmatrix} Q_1 & Q_2 & Q_3 \end{bmatrix}, \quad Q_{12} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \quad Q_{13} = \begin{bmatrix} Q_1 & Q_3 \end{bmatrix} \\ Q_1 &= \begin{bmatrix} \vec{q}_1 & \dots & \vec{q}_q \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \vec{q}_{q+1} & \dots & \vec{q}_r \end{bmatrix}, \quad Q_3 = \begin{bmatrix} \vec{q}_{r+1} & \dots & \vec{q}_n \end{bmatrix}. \end{aligned} \quad (5.1.36)$$

Es sei  $\vec{Z} = (Z_1, \dots, Z_n)^T$  die Koordinatendarstellung von  $\vec{Y}$  bzgl. der kanonischen Basis

$$\vec{Y} = \sum_{i=1}^n \vec{q}_i Z_i = Q\vec{Z}, \quad (5.1.37)$$

dann folgt für  $\vec{Z}$

$$\vec{Z} = Q^T \vec{Y} = \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{bmatrix} X\vec{\beta} + Q^T \vec{\epsilon} = \begin{bmatrix} Q_1^T X\vec{\beta} \\ Q_2^T X\vec{\beta} \\ \vec{0} \end{bmatrix} + Q^T \vec{\epsilon}. \quad (5.1.38)$$

Mit  $\vec{\gamma}_1 := Q_1^T X\vec{\beta}$ ,  $\vec{\gamma}_2 := Q_2^T X\vec{\beta}$  und  $\vec{\gamma} = Q_{12}^T X\vec{\beta} = [\vec{\gamma}_1^T \vec{\gamma}_2^T]^T$ , der eindeutigen Koordinatendarstellung von  $E[\vec{Y}] = X\vec{\beta}$  bezüglich der Basis  $\vec{q}_1, \dots, \vec{q}_r$  des Spaltenraumes von  $X$ , sowie  $\vec{\eta} := Q^T \vec{\epsilon}$  ergibt sich hieraus die kanonische Darstellung des linearen Modells bezüglich der gegebenen kanonischen Basis zu

$$\vec{Z} = \begin{bmatrix} \vec{\gamma} \\ \vec{0}_{n-r} \end{bmatrix} + \vec{\eta}, \quad \vec{\eta} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \vec{\gamma} = (\gamma_1, \dots, \gamma_r)^T \in \mathbb{R}^r \quad (5.1.39)$$

Da  $\vec{\mu} = X\vec{\beta}$  unter der Hypothese  $H$  in  $\mathcal{V}_\omega = \mathcal{V}_{r-q}$ , dem Spaltentaum von  $Q_2$ , liegt, ergibt sich bei Vorliegen von  $H$   $\vec{\gamma}_1 = Q_1^T X\vec{\beta} = \vec{0}$ . Somit lautet das durch die Hypothese  $H$  gegebene reduzierte Modell in dieser kanonischen Darstellung

$$\vec{Z} = \begin{bmatrix} \vec{0}_q \\ \vec{\gamma}_2 \\ \vec{0}_{n-r} \end{bmatrix} + \vec{\eta}, \quad \vec{\eta} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \vec{\gamma}_2 \in \mathbb{R}^{r-q} \quad (5.1.40)$$

Es seien  $\hat{Y}_\Omega = \hat{\mu}_\Omega = X\hat{\beta}_\Omega$  der BLE-Schätzer für  $E[\vec{Y}] = \vec{\mu}_{\vec{Y}} = X\vec{\beta}$  und  $S_\Omega = (\vec{Y} - \hat{Y}_\Omega)^T (\vec{Y} - \hat{Y}_\Omega)$  die Residuenquadratsumme des vollständigen linearen Modells sowie  $\hat{Y}_\omega$  und  $S_\omega$  die entsprechenden Größen des reduzierten Modells.

Mit  $\vec{Z}_1 = (Z_1, \dots, Z_q)^T = Q_1^T \vec{Y}$ ,  $\vec{Z}_2 = (Z_{q+1}, \dots, Z_r)^T = Q_2^T \vec{Y}$  und  $\vec{Z}_3 = (Z_{r+1}, \dots, Z_n)^T = Q_3^T \vec{Y}$  ergibt sich, da die Größen  $\frac{1}{\sigma} Z_1, \dots, \frac{1}{\sigma} Z_n$  unabhängig und normalverteilt mit Varianz 1 sind:

$$\begin{aligned}
 \hat{Y}_\Omega &= p_{\mathcal{V}_\Omega}(\vec{Y}) = P_{\mathfrak{M}(Q_{12})} \vec{Y} = Q_{12} Q_{12}^T \vec{Y} = Q_1 \vec{Z}_1 + Q_2 \vec{Z}_2 = \sum_{i=1}^r \vec{q}_i Z_i \\
 \vec{Y} - \hat{Y}_\Omega &= p_{\mathcal{V}_\Omega^\perp}(\vec{Y}) = P_{\mathfrak{M}(Q_3)} \vec{Y} = Q_3 Q_3^T \vec{Y} = Q_3 \vec{Z}_3 = \sum_{i=r+1}^n \vec{q}_i Z_i \\
 S_\Omega &= (\vec{Y} - \hat{Y}_\Omega)^T (\vec{Y} - \hat{Y}_\Omega) = \vec{Z}_3^T \vec{Z}_3 = \sum_{i=r+1}^n Z_i^2 \\
 \frac{1}{\sigma^2} S_\Omega &= \frac{1}{\sigma^2} \vec{Z}_3^T \vec{Z}_3 = \sum_{i=r+1}^n \left( \frac{Z_i}{\sigma} \right)^2 \sim \chi_{n-r}^2 \\
 \hat{Y}_\omega &= p_{\mathcal{V}_\omega}(\vec{Y}) = P_{\mathfrak{M}(Q_2)} \vec{Y} = Q_2 Q_2^T \vec{Y} = Q_2 \vec{Z}_2 = \sum_{i=q+1}^r \vec{q}_i Z_i \\
 \vec{Y} - \hat{Y}_\omega &= p_{\mathcal{V}_\omega^\perp}(\vec{Y}) = P_{\mathfrak{M}(Q_{13})} \vec{Y} = Q_{13} Q_{13}^T \vec{Y} = Q_1 \vec{Z}_1 + Q_3 \vec{Z}_3 = \sum_{i=1}^q \vec{q}_i Z_i + \sum_{i=r+1}^n \vec{q}_i Z_i \\
 S_\omega &= (\vec{Y} - \hat{Y}_\omega)^T (\vec{Y} - \hat{Y}_\omega) = \vec{Z}_1^T \vec{Z}_1 + \vec{Z}_3^T \vec{Z}_3 = \sum_{i=1}^q Z_i^2 + \sum_{i=r+1}^n Z_i^2 \\
 \hat{Y}_\Omega - \hat{Y}_\omega &= \sum_{i=1}^q \vec{q}_i Z_i = Q_1 \vec{Z}_1 = Q_1 Q_1^T \vec{Y} = P_{\mathfrak{M}(Q_1)} \vec{Y} = p_{\mathcal{V}_{\Omega-\omega}}(\vec{Y}) \\
 \hat{Y}_\Omega - \hat{Y}_\omega &= P_{\mathfrak{M}(Q_1)} (\hat{Y}_\Omega + \vec{Y} - \hat{Y}_\Omega) = P_{\mathfrak{M}(Q_1)} \hat{Y}_\Omega = p_{\mathcal{V}_{\Omega-\omega}}(\hat{Y}_\Omega) \\
 (\hat{Y}_\Omega - \hat{Y}_\omega)^T (\hat{Y}_\Omega - \hat{Y}_\omega) &= \vec{Z}_1^T \vec{Z}_1 = S_\omega - S_\Omega = \sum_{i=1}^q Z_i^2 \\
 \frac{1}{\sigma^2} (S_\omega - S_\Omega) &= \frac{1}{\sigma^2} \vec{Z}_1^T \vec{Z}_1 = \sum_{i=1}^q \left( \frac{Z_i}{\sigma} \right)^2 \sim \chi_{q,\delta}^2 \\
 \delta &= \frac{1}{\sigma^2} (\gamma_1^2 + \dots + \gamma_q^2)
 \end{aligned} \tag{5.1.41}$$

Aus der Unabhängigkeit der Zufallsvektoren  $\vec{Z}_1$ ,  $\vec{Z}_2$  und  $\vec{Z}_3$  folgt die Unabhängigkeit von  $S_\Omega$  und  $S_\omega - S_\Omega$ . Somit ist

$$F = \frac{\frac{1}{q} (S_\omega - S_\Omega)}{\frac{1}{n-r} S_\Omega} = \frac{\frac{1}{q\sigma^2} (S_\omega - S_\Omega)}{\frac{1}{(n-r)\sigma^2} S_\Omega} \sim F_{q,n-r,\delta} \tag{5.1.42}$$

mit  $\delta = 0$  unter  $H$ .

### Beweis:

Die kanonische Darstellung für das vollständige lineare Modell wurde in Satz 2.2.2 sowie im Beweis des Satzes 3.3.2 hergeleitet. Die orthogonalen Projektionen ergeben sich mit Korollar 1.2 zu Definition 1.2.3.

**Satz 5.1.4.**

Es sei  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  eine  $q$ -dimensionale schätzbare Funktion mit  $rg(C) = q$  des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells mit  $rg(X) = r$  ( $r \leq k$ ),  $\hat{\psi} = A\vec{Y} = C\hat{\beta} = AX\hat{\beta}$  der BLE-Schätzer von  $\vec{\psi}$ ,  $\hat{\beta} = \hat{\beta}_\Omega$  eine beliebige Lösung der Normalgleichungen und  $B = AA^T = C(X^T X)^{-1}C^T$ .

## 1. Für die testbare Hypothese

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{0} \quad (5.1.43)$$

ergibt sich mit Gl. (5.1.41)

$$\begin{aligned} \hat{Y}_\Omega - \hat{Y}_\omega &= p_{\mathcal{V}_{\Omega-\omega}}(\hat{Y}_\Omega) = P_{\mathfrak{M}(A^T)}\hat{Y}_\Omega = A^T(AA^T)^{-1}A\hat{Y}_\Omega \\ S_\omega - S_\Omega &= (\hat{Y}_\Omega - \hat{Y}_\omega)^T(\hat{Y}_\Omega - \hat{Y}_\omega) \\ &= \hat{Y}_\Omega^T A^T(AA^T)^{-1}AA^T(AA^T)^{-1}A\hat{Y}_\Omega \\ &= \hat{Y}_\Omega^T A^T(AA^T)^{-1}A\hat{Y}_\Omega \\ &= \hat{\psi}^T B^{-1}\hat{\psi} = (C\hat{\beta})^T B^{-1}C\hat{\beta} \end{aligned} \quad (5.1.44)$$

## 2. Für die testbare Hypothese

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{c} \quad (5.1.45)$$

ergibt sich unter Verwendung des transformierten Modells aus Bemerkung 2 zu Definition 5.1.1 und Gl. (5.1.44)

$$\begin{aligned} S_\omega - S_\Omega &= S_{\omega_{H^*}}^* - S_\Omega^* = (C\hat{\beta}^*)^T B^{-1}C\hat{\beta}^* = (C(\hat{\beta} - \tilde{\beta}))^T B^{-1}C(\hat{\beta} - \tilde{\beta}) \\ &= (C\hat{\beta} - \vec{c})^T B^{-1}(C\hat{\beta} - \vec{c}) = (\hat{\psi} - \vec{c})^T B^{-1}(\hat{\psi} - \vec{c}) \end{aligned} \quad (5.1.46)$$

sowie

$$\begin{aligned} F &= \frac{\frac{1}{q}\|\hat{Y}_\Omega - \hat{Y}_\omega\|^2}{\frac{1}{n-r}S_\Omega} = \frac{\frac{1}{q}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} = \frac{\frac{1}{q}(S_{\omega_{H^*}}^* - S_\Omega^*)}{\frac{1}{n-r}S_\Omega^*} \\ &= \frac{(C\hat{\beta} - \vec{c})^T B^{-1}(C\hat{\beta} - \vec{c})}{q\hat{\sigma}^2} = \frac{\frac{1}{q\hat{\sigma}^2}(\hat{\psi} - \vec{c})^T B^{-1}(\hat{\psi} - \vec{c})}{\frac{1}{(n-r)\hat{\sigma}^2}(\vec{Y} - X\hat{\beta})^T(\vec{Y} - X\hat{\beta})} \\ &\sim F_{q, n-r, \delta}, \quad \delta = \frac{1}{\hat{\sigma}^2}(C\vec{\beta} - \vec{c})^T B^{-1}(C\vec{\beta} - \vec{c}) \end{aligned} \quad (5.1.47)$$

Wenn  $H$  vorliegt, ist  $\delta = 0$ .

**Satz 5.1.5. F-Test**

Es sei  $\hat{\psi} = C\hat{\beta} = A\vec{Y}$  der BLE-Schätzer für die  $q$ -dimensionale schätzbare Funktion  $\vec{\psi}(\vec{\beta}) = C\vec{\beta}$  des Parametervektors  $\vec{\beta}$  des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells mit  $\text{rg}(C) = q$  und  $B = AA^T = C(X^T X)^- C^T$ . Für die testbare Hypothese über die Parameter des klassischen linearen Modells

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{c} \quad (5.1.48)$$

lautet der Likelihood-Quotiententest zum Niveau  $\alpha$  wie folgt. Die Hypothese wird abgelehnt genau dann, wenn

$$F = \frac{\frac{1}{q}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} = \frac{(C\hat{\beta} - \vec{c})^T B^{-1}(C\hat{\beta} - \vec{c})}{q\hat{\sigma}^2} > F_{q, n-r; 1-\alpha} \quad (5.1.49)$$

## 5.2 Tests spezieller Hypothesen

### Satz 5.2.1. t-Test für eine eindimensionale schätzbare Funktion

Es sei  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta}$  eine eindimensionale schätzbare Funktion des Parametervektors  $\vec{\beta}$  des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells und  $\hat{\psi} = \vec{c}^T \hat{\beta}$  der BLE-Schätzer von  $\psi$  mit  $\sigma_{\hat{\psi}}^2 = \sigma^2 B$  mit  $B = \vec{c}^T (X^T X)^{-1} \vec{c}$ . Für die Testgröße des F-Tests der Hypothese

$$H : \quad \psi(\vec{\beta}) = \vec{c}^T \vec{\beta} = c \quad (5.2.1)$$

ergibt sich

$$F = \frac{(\vec{c}^T \hat{\beta} - c)^T B^{-1} (\vec{c}^T \hat{\beta} - c)}{\hat{\sigma}^2} = \frac{(\vec{c}^T \hat{\beta} - c)^2}{\hat{\sigma}^2 B} = \left( \frac{\hat{\psi} - c}{\hat{\sigma}_{\hat{\psi}}} \right)^2 \quad (5.2.2)$$

Aus

$$\frac{\hat{\psi} - c}{\sigma_{\hat{\psi}}} \sim \mathfrak{N}(\delta, 1), \quad \delta = \frac{\psi(\vec{\beta}) - c}{\sigma_{\hat{\psi}}} \quad (5.2.3)$$

$$(n - r) \frac{\hat{\sigma}_{\hat{\psi}}^2}{\sigma_{\hat{\psi}}^2} = (n - r) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2 \quad (5.2.4)$$

folgt, da  $\hat{\psi}$  und  $\hat{\sigma}^2$  stochastisch unabhängig sind,

$$t = \frac{\hat{\psi} - c}{\hat{\sigma}_{\hat{\psi}}} = \frac{\frac{\hat{\psi} - c}{\sigma_{\hat{\psi}}}}{\sqrt{\frac{\hat{\sigma}_{\hat{\psi}}^2}{\sigma_{\hat{\psi}}^2}}} \sim t_{n-r, \delta} \quad (5.2.5)$$

mit  $\delta = 0$  unter  $H$ . Da außerdem  $t^2 = F$ , ergibt sich

$$t_{n-r; 1-\frac{\alpha}{2}} = \sqrt{F_{1, n-r; 1-\alpha}} \quad (5.2.6)$$

Somit kann der F-Test für die Hypothese  $H$  wie folgt formuliert werden. Die Hypothese  $H$  wird zum Niveau  $\alpha$  abgelehnt genau dann, wenn

$$|t| = \left| \frac{\hat{\psi} - c}{\hat{\sigma}_{\hat{\psi}}} \right| > t_{n-r; 1-\frac{\alpha}{2}} \quad (5.2.7)$$

Der F-Test für die Hypothese  $H$  ist äquivalent zu dem zweiseitigen t-Test der Hypothese  $H$  gegen die Alternative  $A : \psi(\vec{\beta}) = \vec{c}^T \vec{\beta} \neq c$ .  $H$  wird zum Niveau  $\alpha$  zugunsten der Alternative  $A : \psi(\vec{\beta}) = \vec{c}^T \vec{\beta} > c$  abgelehnt genau dann, wenn  $t > t_{n-r; 1-\alpha}$  und zugunsten der Alternative  $A : \psi(\vec{\beta}) = \vec{c}^T \vec{\beta} < c$  abgelehnt genau dann, wenn  $t < -t_{n-r; 1-\alpha}$ .

**Beispiel 5.2.1. t-Test für einen Parameter bei vollem Spaltenrang**

Die Designmatrix des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells besitze vollen Spaltenrang. Dann ist  $\beta_i$ , die  $i$ -te Komponente des Parametervektors  $\vec{\beta}$ , schätzbar. Aus Satz 5.2.1 ergeben sich mit  $\vec{c} = \vec{e}_i$  ( $\vec{e}_i$  sei der  $i$ -te Einheitsvektor),  $\hat{\psi} = \hat{\beta}_i$ , dem BLE-Schätzer von  $\psi(\vec{\beta}) = \vec{c}^T \vec{\beta} = \beta_i$ , und  $B = d_i$ , dem  $i$ -ten Diagonalelement von  $(X^T X)^{-1}$ , die folgenden t-Tests für die Hypothese

$$H : \beta_i = \beta_{i,0} . \quad (5.2.8)$$

Die Hypothese  $H$  wird zum Niveau  $\alpha$  zugunsten der Alternative  $A : \beta_i \neq \beta_{i,0}$  abgelehnt genau dann, wenn

$$|t| = \left| \frac{\hat{\beta}_i - \beta_{i,0}}{\hat{\sigma} \sqrt{d_i}} \right| > t_{n-k; 1-\frac{\alpha}{2}} \quad (5.2.9)$$

$H$  wird zum Niveau  $\alpha$  zugunsten der Alternative  $A : \beta_i > \beta_{i,0}$  abgelehnt genau dann, wenn  $t > t_{n-k; 1-\alpha}$  und zugunsten der Alternative  $A : \beta_i < \beta_{i,0}$  abgelehnt genau dann, wenn  $t < -t_{n-k; 1-\alpha}$ .

Die für die Berechnung der Testgröße  $F = t^2$  mittels der Residuenquadratsummen  $S_\Omega$  und  $S_\omega$  gemäß Bemerkung 2 zu Definition 5.1.1 erforderlichen Transformationen ergeben sich hier zu

$$\vec{\beta}^* = \vec{\beta} - \beta_{i,0} \vec{e}_i , \quad \vec{Y}^* = \vec{Y} - \beta_{i,0} \vec{x}_i \quad (5.2.10)$$

wobei  $\vec{x}_i$  der  $i$ -te Spaltenvektor der Designmatrix  $X$  ist.

Der Parametervektor  $\vec{\beta}$  des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells erfüllt die Hypothese  $H$  genau dann, wenn der Parametervektor  $\vec{\beta}^*$  des klassischen linearen Modells

$$\vec{Y}^* = X \vec{\beta}^* + \vec{\epsilon} , \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I) , \quad rg(X) = r < n , \quad \begin{pmatrix} \vec{\beta}^* \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^k \times \mathbb{R}_{>0} , \quad (5.2.11)$$

die Hypothese

$$H : \beta_i^* = 0 \quad (5.2.12)$$

erfüllt, wobei  $\beta_i^*$  die  $i$ -te Komponente des Parametervektors  $\vec{\beta}^*$  ist.

**Beispiel 5.2.2. Hypothesentests für die Parameter der einfachen linearen Regression**

Für die Parameter des durch die Gleichungen

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i , \quad \epsilon_i \sim \mathfrak{N}(0, \sigma^2) \quad (i = 1, \dots, n) , \quad \sigma^2 > 0 \\ &= \alpha + \beta(x_i - \bar{x}) + \epsilon_i \\ \epsilon_1, \dots, \epsilon_n &\text{ stochastisch unabhängig} \end{aligned} \quad (5.2.13)$$

gegebenen einfachen linearen Regressionsmodells ergeben sich mit Beispiel 5.2.1 die folgenden Testverfahren.

Die Testgröße des  $t$ -Tests ist

- für die Hypothese  $H : \beta_0 = \beta_{0,0} : t = \frac{\hat{\beta}_0 - \beta_{0,0}}{\hat{\sigma} \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}}$
- für die Hypothese  $H : \beta_1 = \beta_{1,0} : t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\hat{\sigma} \sqrt{\frac{1}{n s_{xx}}}}$
- für die Hypothese  $H : \alpha = \alpha_0 : t = \frac{\hat{\alpha} - \alpha_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$

Die Hypothese  $H : \theta = \theta_0$  ( $\theta = \beta_0, \beta_1, \alpha; \theta_0 = \beta_{0,0}, \beta_{1,0}, \alpha_0$ ) wird zum Niveau  $\alpha'$  abgelehnt zugunsten

- der Alternative  $A : \theta \neq \theta_0$  genau dann, wenn  $|t| > t_{n-2; 1-\frac{\alpha'}{2}}$
- der Alternative  $A : \theta > \theta_0$  genau dann, wenn  $t > t_{n-2; 1-\alpha'}$
- der Alternative  $A : \theta < \theta_0$  genau dann, wenn  $t < -t_{n-2; 1-\alpha'}$

Satz 5.1.5 ergibt für die Hypothese  $H : \beta_0 = \beta_{0,0}, \beta_1 = \beta_{1,0}$  mit  $C = I, \vec{\beta} = (\beta_0, \beta_1)^T, \vec{c} = (\beta_{0,0}, \beta_{1,0})^T$  und

$$B^{-1} = X^T X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \quad (5.2.14)$$

den folgenden  $F$ -Test. Die Hypothese  $H$  wird zum Niveau  $\alpha'$  abgelehnt genau dann, wenn

$$F = \frac{(C\hat{\beta} - \vec{c})^T B^{-1} (C\hat{\beta} - \vec{c})}{2\hat{\sigma}^2} \quad (5.2.15)$$

$$= \frac{n(\hat{\beta}_0 - \beta_{0,0})^2 + 2(\sum_{i=1}^n x_i)(\hat{\beta}_0 - \beta_{0,0})(\hat{\beta}_1 - \beta_{1,0}) + (\sum_{i=1}^n x_i^2)(\hat{\beta}_1 - \beta_{1,0})^2}{2\hat{\sigma}^2} > F_{2, n-2; 1-\alpha'} \quad (5.2.16)$$

### Beispiel 5.2.3. F-Test für den Parametervektor bei vollem Spaltenrang

Es sei  $\hat{\beta}$  der BLE-Schätzer für den Parametervektor des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells, dessen Designmatrix  $X$  vollen Spaltenrang besitzt ( $\text{rg}(X) = k$ ). Der  $F$ -Test nach Satz 5.1.5 für die Hypothese

$$H : \vec{\beta} = \vec{\beta}_0 \quad (5.2.17)$$

zum Niveau  $\alpha$  führt, da  $B^{-1} = X^T X$ , genau dann zur Ablehnung von  $H$ , wenn

$$F = \frac{(\hat{\beta} - \vec{\beta}_0)^T X^T X (\hat{\beta} - \vec{\beta}_0)}{k\hat{\sigma}^2} > F_{k, n-k; 1-\alpha} . \quad (5.2.18)$$

### Beispiel 5.2.4. F-Test auf Signifikanz des Modells

Mit dem  $F$ -Test auf Signifikanz des Modells, der auch Test auf Gesamtzusammenhang genannt wird, wird untersucht, ob die Einflussgrößen des Modells zusammen einen signifikanten Einfluß auf die Zielgröße ausüben.



1. Für ein Modell ohne Achsenabschnitt wird die Hypothese

$$H : \vec{\mu}_{\vec{Y}} = X\vec{\beta} = \vec{0} \quad (5.2.19)$$

getestet. Es sei  $X^{(r)}$  eine Matrix, die aus  $r$  linear unabhängigen Zeilen der Designmatrix  $X$  besteht, dann ist die Hypothese  $H$  äquivalent zu der Hypothese  $H^{(r)} : X^{(r)}\vec{\beta} = \vec{0}$ .

Besitzt die Designmatrix vollen Spaltenrang ( $\text{rg}(X) = k$ ) dann lautet die Hypothese

$$H : \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0 \quad (5.2.20)$$

Das reduzierte Modell unter  $H$  ist

$$\vec{Y} = \vec{\mu} + \vec{\epsilon} = \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \vec{\mu} \in \mathcal{V}_\omega = \mathcal{V}_{r-q} = \{\vec{0}\}, \quad r - q = 0 \quad (5.2.21)$$

Somit ist  $\hat{Y}_\omega = \vec{0}$  und  $S_\omega - S_\Omega = (\hat{Y}_\Omega - \hat{Y}_\omega)^T (\hat{Y}_\Omega - \hat{Y}_\omega) = \hat{Y}_\Omega^T \hat{Y}_\Omega = SQ_R$  die Modellquadratsumme des vollständigen linearen Modells. Der  $F$ -Test für  $H$  zum Niveau  $\alpha$  führt zur Ablehnung von  $H$  genau dann, wenn

$$F = \frac{\frac{1}{r}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} = \frac{SQ_R}{r\hat{\sigma}^2} = \frac{\hat{\beta}^T X^T X \hat{\beta}}{r\hat{\sigma}^2} > F_{r,n-r;1-\alpha} \quad (5.2.22)$$

wobei  $\hat{\beta} = \hat{\beta}_\Omega$  eine Lösung der Normalgleichungen des vollständigen linearen Modells ist.

2. Für das klassische lineare Modell mit Achsenabschnitt für den  $n$ -dimensionalen Beobachtungsvektor  $\vec{Y} = (Y_1, \dots, Y_n)^T$

$$\begin{aligned} Y_i &= \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k + \epsilon_i, \quad i = 1, \dots, n \\ \vec{Y} &= X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X) = r < n, \\ \vec{\beta} &= (\beta_0, \beta_1, \dots, \beta_k)^T \in \mathbb{R}^{k+1}, \quad \begin{pmatrix} \vec{\beta} \\ \sigma^2 \end{pmatrix} \in \Omega = \mathbb{R}^{k+1} \times \mathbb{R}_{>0} \end{aligned} \quad (5.2.23)$$

wird die durch das reduzierte Modell

$$\begin{aligned} Y_i &= \mu + \epsilon_i, \quad i = 1, \dots, n \\ \vec{Y} &= \vec{1}_n \mu + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \mu \in \mathbb{R} \end{aligned} \quad (5.2.24)$$

gegebene Hypothese  $H$  getestet. Besitzt die Designmatrix vollen Spaltenrang ( $\text{rg}(X) = k + 1$ ) dann lautet die Hypothese

$$H : \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0 \quad (5.2.25)$$

Somit ist  $r - q = 1$ ,  $\hat{Y}_\omega = \vec{1}_n \bar{Y}$  und  $S_\omega - S_\Omega = (\hat{Y}_\Omega - \hat{Y}_\omega)^T (\hat{Y}_\Omega - \hat{Y}_\omega) = (\hat{Y}_\Omega - \vec{1}_n \bar{Y})^T (\hat{Y}_\Omega - \vec{1}_n \bar{Y}) = SQ_{Rc}$  die um den Mittelwert bereinigte Modellquadratsumme des vollständigen linearen Modells. Der  $F$ -Test für  $H$  führt zur Ablehnung von  $H$  genau dann, wenn

$$F = \frac{\frac{1}{r-1}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} = \frac{SQ_{cR}}{(r-1)\hat{\sigma}^2} > F_{r-1,n-r;1-\alpha}. \quad (5.2.26)$$

**Bemerkung 1. Streuungszerlegung**

Es seien im Folgenden  $H$  eine testbare Hypothese der Form

$$H : \quad \vec{\psi}(\vec{\beta}) = C\vec{\beta} = \vec{0} \quad (5.2.27)$$

für das durch Gleichung (5.1.1) gegebene lineare Modell,  $\vec{q}_1, \dots, \vec{q}_n$  eine kanonische Basis für die testbare Hypothese und  $Z_1, \dots, Z_n$  die kanonischen Koordinaten des Beobachtungsvektors.

1. Die Zerlegung des Beobachtungsvektors  $\vec{Y} \in \mathbb{R}^n$  (siehe auch Bemerkung 1 in Abschnitt 2.2)

$$\vec{Y} = \sum_{i=1}^n \vec{q}_i Z_i = \hat{Y}_\Omega + (\vec{Y} - \hat{Y}_\Omega) = \hat{Y}_\omega + (\hat{Y}_\Omega - \hat{Y}_\omega) + (\vec{Y} - \hat{Y}_\Omega) \quad (5.2.28)$$

in orthogonale Komponenten, wobei

$$\begin{aligned} \hat{Y}_\Omega &= \sum_{i=1}^r \vec{q}_i Z_i \in \mathcal{V}_\Omega = \mathcal{V}_r, & \vec{Y} - \hat{Y}_\Omega &= \sum_{i=r+1}^n \vec{q}_i Z_i \in \mathcal{V}_\Omega^\perp = \mathcal{V}_r^\perp \\ \hat{Y}_\omega &= \sum_{i=q+1}^r \vec{q}_i Z_i \in \mathcal{V}_\omega = \mathcal{V}_{r-q}, & \hat{Y}_\Omega - \hat{Y}_\omega &= \sum_{i=1}^q \vec{q}_i Z_i \in \mathcal{V}_{\Omega-\omega} = \mathcal{V}_q \end{aligned} \quad (5.2.29)$$

ergibt mit (siehe Gleichungen (5.1.41))

$$\begin{aligned} SQ_T &= \vec{Y}^T \vec{Y} = \sum_{i=1}^n Z_i^2 \\ SQ_E &= (\vec{Y} - \hat{Y}_\Omega)^T (\vec{Y} - \hat{Y}_\Omega) = S_\Omega = \sum_{i=r+1}^n Z_i^2 \\ SQ_R &= \hat{Y}_\Omega^T \hat{Y}_\Omega = \sum_{i=1}^r Z_i^2 \\ SQ_{R,\omega} &= \hat{Y}_\omega^T \hat{Y}_\omega = \sum_{i=q+1}^r Z_i^2 \\ SQ_{R,H} &= (\hat{Y}_\Omega - \hat{Y}_\omega)^T (\hat{Y}_\Omega - \hat{Y}_\omega) = S_\omega - S_\Omega = \sum_{i=1}^q Z_i^2 \end{aligned} \quad (5.2.30)$$

unter Anwendung des Satzes von Pythagoras die folgenden Zerlegungen der Gesamtstreuung und der Modellquadratsumme

$$\begin{aligned} SQ_T &= SQ_R + SQ_E = SQ_{R,\omega} + SQ_{R,H} + SQ_E \\ SQ_R &= SQ_{R,\omega} + SQ_{R,H} \end{aligned} \quad (5.2.31)$$

Zusätzlich zu der durch das lineare Modell gegebenen Zerlegung der Gesamtstreuung in eine Modellquadratsumme  $SQ_R$  und eine Fehlerquadratsumme  $SQ_E$  ergibt die Hypothese  $H$  eine Zerlegung der Modellquadratsumme  $SQ_R$  des vollständigen linearen Modells in die Modellquadratsumme des reduzierten Modells  $SQ_{R,\omega}$  und die Quadratsumme  $SQ_{R,H}$ , die auf die Erweiterung des reduzierten Modells zum vollständigen Modell zurückzuführen ist.

2. Für Modelle mit Achsenabschnitt sowohl im vollständigen als auch im reduzierten Modell gilt

$$\begin{aligned}\vec{1}_n &\in \mathcal{V}_r = \mathcal{L}(\vec{q}_1, \dots, \vec{q}_r) = \mathfrak{Im}(X) = \{\vec{\mu} = X\vec{\beta} \mid (\vec{\beta}^T, \sigma^2)^T \in \Omega\} \\ \vec{1}_n &\in \mathcal{V}_{r-q} = \mathcal{L}(\vec{q}_{q+1}, \dots, \vec{q}_r) = \{\vec{\mu} = X\vec{\beta} \mid (\vec{\beta}^T, \sigma^2)^T \in \omega\}\end{aligned}\quad (5.2.32)$$

Wir wählen die kanonische Basis daher so, dass  $\vec{q}_r = \frac{1}{\sqrt{n}}\vec{1}_n$ , dann gilt  $Z_r = \sqrt{n}\bar{Y}$ .

Die Zerlegung des Beobachtungsvektors  $\vec{Y} \in \mathbb{R}^n$  (siehe auch Bemerkung 3 in Abschnitt 2.2)

$$\begin{aligned}\vec{Y} &= \sum_{i=1}^n \vec{q}_i Z_i = \bar{Y}\vec{1}_n + (\vec{Y} - \bar{Y}\vec{1}_n) = \bar{Y}\vec{1}_n + (\hat{Y}_\Omega - \bar{Y}\vec{1}_n) + (\vec{Y} - \hat{Y}_\Omega) \\ &= \bar{Y}\vec{1}_n + (\hat{Y}_\omega - \bar{Y}\vec{1}_n) + (\hat{Y}_\Omega - \hat{Y}_\omega) + (\vec{Y} - \hat{Y}_\Omega)\end{aligned}\quad (5.2.33)$$

in orthogonale Komponenten, wobei

$$\begin{aligned}\bar{Y}\vec{1}_n &= q_r Z_r \in \mathcal{L}(\vec{q}_r), \quad \vec{Y} - \bar{Y}\vec{1}_n = \sum_{i=1}^{r-1} \vec{q}_i Z_i + \sum_{i=r+1}^n \vec{q}_i Z_i \in \mathcal{L}(\vec{q}_r)^\perp \\ \hat{Y}_\Omega - \bar{Y}\vec{1}_n &= \sum_{i=1}^{r-1} \vec{q}_i Z_i \in \mathcal{L}(\vec{q}_1, \dots, \vec{q}_{r-1}), \quad \vec{Y} - \hat{Y}_\Omega = \sum_{i=r+1}^n \vec{q}_i Z_i \in \mathcal{V}_\Omega^\perp = \mathcal{V}_r^\perp \\ \hat{Y}_\omega - \bar{Y}\vec{1}_n &= \sum_{i=q+1}^{r-1} \vec{q}_i Z_i \in \mathcal{L}(\vec{q}_{q+1}, \dots, \vec{q}_{r-1}), \quad \hat{Y}_\Omega - \hat{Y}_\omega = \sum_{i=1}^q \vec{q}_i Z_i \in \mathcal{V}_{\Omega-\omega} = \mathcal{V}_q\end{aligned}\quad (5.2.34)$$

ergibt mit

$$\begin{aligned}SQ_T &= \vec{Y}^T \vec{Y} = \sum_{i=1}^n Z_i^2, \quad SQ_M = n\bar{Y}^2 = Z_r^2 \\ SQ_{Tc} &= (\vec{Y} - \bar{Y}\vec{1}_n)^T (\vec{Y} - \bar{Y}\vec{1}_n) = \sum_{i=1}^{r-1} Z_i^2 + \sum_{i=r+1}^n Z_i^2 = SQ_T - SQ_M \\ SQ_E &= (\vec{Y} - \hat{Y}_\Omega)^T (\vec{Y} - \hat{Y}_\Omega) = S_\Omega = \sum_{i=r+1}^n Z_i^2 \\ SQ_{Rc} &= (\hat{Y}_\Omega - \bar{Y}\vec{1}_n)^T (\hat{Y}_\Omega - \bar{Y}\vec{1}_n) = \sum_{i=1}^{r-1} Z_i^2 = SQ_R - SQ_M \\ SQ_{Rc,\omega} &= (\hat{Y}_\omega - \bar{Y}\vec{1}_n)^T (\hat{Y}_\omega - \bar{Y}\vec{1}_n) = \sum_{i=q+1}^{r-1} Z_i^2 = SQ_{R,\omega} - SQ_M \\ SQ_{R,H} &= (\hat{Y}_\Omega - \hat{Y}_\omega)^T (\hat{Y}_\Omega - \hat{Y}_\omega) = S_\omega - S_\Omega = \sum_{i=1}^q Z_i^2\end{aligned}\quad (5.2.35)$$

unter Anwendung des Satzes von Pythagoras die folgenden Zerlegungen der unbereinigten und

der um den Mittelwert bereinigten Gesamtstreuung sowie Modellquadratsumme

$$\begin{aligned}
 SQ_T &= SQ_M + SQ_{Rc} + SQ_E = SQ_M + SQ_{Rc,\omega} + SQ_{R,H} + SQ_E \\
 SQ_{Tc} &= SQ_{Rc} + SQ_E = SQ_{Rc,\omega} + SQ_{R,H} + SQ_E \\
 SQ_R &= SQ_M + SQ_{cR,\omega} + SQ_{R,H} \\
 SQ_{Rc} &= SQ_{Rc,\omega} + SQ_{R,H}
 \end{aligned} \tag{5.2.36}$$

Zusätzlich zu der durch das lineare Modell gegebenen Zerlegung der bereinigten Gesamtstreuung in eine bereinigte Modellquadratsumme  $SQ_{Rc}$  und eine Fehlerquadratsumme  $SQ_E$  ergibt die Hypothese  $H$  eine Zerlegung der bereinigten Modellquadratsumme  $SQ_{Rc}$  des vollständigen linearen Modells in die bereinigte Modellquadratsumme des reduzierten Modells  $SQ_{Rc,\omega}$  und die Quadratsumme  $SQ_{R,H}$ , die auf die Erweiterung des reduzierten Modells zum vollständigen Modell zurückzuführen ist.

3. Jede orthogonale Komponente variiert in einem Vektorraum und die zugehörige Quadratsumme, das Quadrat der Norm der Komponente, ist darstellbar als Summe der Quadrate unabhängiger normalverteilter Zufallsgrößen. Die Anzahl der Summanden, die übereinstimmt mit der Dimension des zugehörigen Vektorraums, wird als Anzahl der Freiheitsgrade der Quadratsumme bzw. der Komponente bezeichnet.

Tabelle der Streuungszerlegung (Gesamtstreuung):

Ursache	FG	SQ	MQ	F-Wert
Modell	$r$	$SQ_R = \hat{Y}_\Omega^T \hat{Y}_\Omega$	$MQ_R = \frac{1}{r} SQ_R$	$\frac{MQ_R}{MQ_E}$
red. Modell	$r - q$	$SQ_{R,\omega} = \hat{Y}_\omega^T \hat{Y}_\omega$	$MQ_{R,\omega} = \frac{1}{r-q} SQ_{R,\omega}$	$\frac{MQ_{R,H}}{MQ_E}$
Hypothese $H$	$q$	$SQ_{R,H} = (\hat{Y}_\Omega - \hat{Y}_\omega)^T (\hat{Y}_\Omega - \hat{Y}_\omega)$	$MQ_{R,H} = \frac{1}{q} SQ_{R,H}$	
Fehler	$n - r$	$SQ_E = (\vec{Y} - \hat{Y}_\Omega)^T (\vec{Y} - \hat{Y}_\Omega)$	$MQ_E = \frac{1}{n-r} SQ_E$	
Total	$n$	$SQ_T = \vec{Y}^T \vec{Y}$		

Tabelle der Streuungszerlegung (Bereinigte Gesamtstreuung bei Achsenabschnittsmodell):

Ursache	FG	SQ	MQ	F-Wert
Modell	$r - 1$	$SQ_{Rc} = \ \hat{Y}_\Omega - \bar{Y} \vec{1}_n\ ^2$	$MQ_{Rc} = \frac{1}{r-1} SQ_{Rc}$	$\frac{MQ_{Rc}}{MQ_E}$
red. Modell	$r - 1 - q$	$SQ_{Rc,\omega} = \ \hat{Y}_\omega - \bar{Y} \vec{1}_n\ ^2$	$MQ_{Rc,\omega} = \frac{1}{r-1-q} SQ_{Rc,\omega}$	$\frac{MQ_{R,H}}{MQ_E}$
Hypothese $H$	$q$	$SQ_{R,H} = \ \hat{Y}_\Omega - \hat{Y}_\omega\ ^2$	$MQ_{R,H} = \frac{1}{q} SQ_{R,H}$	
Fehler	$n - r$	$SQ_E = \ \vec{Y} - \hat{Y}_\Omega\ ^2$	$MQ_E = \frac{1}{n-r} SQ_E$	
Korr. Total	$n - 1$	$SQ_{Tc} = \ \vec{Y} - \bar{Y} \vec{1}_n\ ^2$		

Analog zu  $SQ_R = \hat{Y}_\Omega^T \hat{Y}_\Omega = \hat{\beta}_\Omega^T X^T \vec{Y}$  für das vollständige Modell ergibt sich  $SQ_{R,\omega} = \hat{Y}_\omega^T \hat{Y}_\omega = \hat{\beta}_\omega^T X^T \vec{Y}$  für das reduzierte Modell, da  $\hat{Y}_\omega \in \mathcal{V}_{r-q}$  und  $\vec{Y} - \hat{Y}_\omega \in \mathcal{V}_{r-q}^\perp$ .

Für die Testgröße des  $F$ -Tests für die Hypothese  $H$  ergibt sich

$$F = \frac{\frac{1}{q}(S_\omega - S_\Omega)}{\frac{1}{n-r}S_\Omega} = \frac{\frac{1}{q}SQ_{R,H}}{\frac{1}{n-r}S_\Omega} = \frac{MQ_{R,H}}{MQ_E} \quad (5.2.37)$$

und für die Testgrößen der  $F$ -Tests auf Signifikanz des Modells in Gleichung (5.2.22)

$$F = \frac{SQ_R}{r\hat{\sigma}^2} = \frac{MQ_R}{MQ_E} \quad (5.2.38)$$

sowie in Gleichung (5.2.26)

$$F = \frac{SQ_{Rc}}{(r-1)\hat{\sigma}^2} = \frac{MQ_{Rc}}{MQ_E} \quad (5.2.39)$$

### Beispiel 5.2.5. Hypothesentest für mehrere Parameter bei vollem Spaltenrang

Die Designmatrix des durch Gleichung (5.1.1) gegebenen klassischen linearen Modells besitze vollen Spaltenrang. Dann ist der Parametervektor  $\vec{\beta}$  und somit auch jeder Teilvektor schätzbar. Mit den Partitionierungen

$$\vec{\beta} = \begin{bmatrix} \vec{\beta}_1 \\ \vec{\beta}_2 \end{bmatrix}, \quad \hat{\beta} = \hat{\beta}_\Omega = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}, \quad \vec{\beta}_1 = (\beta_1, \dots, \beta_p)^T \quad (5.2.40)$$

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}, \quad X^T X = \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix}, \quad X_1 \in \mathbb{R}^{n \times p}, \quad M_{ij} = X_i^T X_j$$

$$X^T X = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \quad (X^T X)^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{bmatrix}, \quad M_{11}, M^{1,1} \in \mathbb{R}^{p \times p}$$

ergibt sich das vollständige Modell zu

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} = X_1\vec{\beta}_1 + X_2\vec{\beta}_2 + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad rg(X) = k < n, \quad \vec{\beta} \in \mathbb{R}^k \quad (5.2.41)$$

#### 1. Der $F$ -Test für die Hypothese

$$H : \vec{\beta}_2 = \vec{\beta}_{2,0} \quad (5.2.42)$$

führt genau dann zur Ablehnung von  $H$  zum Niveau  $\alpha$ , wenn

$$F = \frac{(\hat{\beta}_2 - \vec{\beta}_{2,0})^T (M^{22})^{-1} (\hat{\beta}_2 - \vec{\beta}_{2,0})}{(k-p)\hat{\sigma}^2} > F_{k-p, n-k; 1-\alpha} \quad (5.2.43)$$

#### 2. Das durch die Hypothese

$$H : \beta_{p+1} = 0, \dots, \beta_k = 0 \quad (5.2.44)$$

beziehungsweise

$$H : \vec{\beta}_2 = \vec{0} \quad (5.2.45)$$

gegebene reduzierte Modell lautet

$$\vec{Y} = X_1 \vec{\beta}_1 + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X_1) = p < k, \quad \vec{\beta}_1 \in \mathbb{R}^p \quad (5.2.46)$$

Der  $F$ -Test zum Niveau  $\alpha$  führt zur Ablehnung der Hypothese  $H$  genau dann, wenn

$$F = \frac{\frac{1}{k-p}(S_\omega - S_\Omega)}{\frac{1}{n-k}S_\Omega} = \frac{\hat{\beta}_2^T (M^{22})^{-1} \hat{\beta}_2}{(k-p)\hat{\sigma}^2} > F_{k-p, n-k; 1-\alpha} \quad (5.2.47)$$

Die Normalgleichungen für das vollständige Modell lauten

$$\begin{aligned} X_1^T X_1 \hat{\beta}_1 + X_1^T X_2 \hat{\beta}_2 &= X_1^T \vec{Y} \\ X_2^T X_1 \hat{\beta}_1 + X_2^T X_2 \hat{\beta}_2 &= X_2^T \vec{Y} \end{aligned} \quad (5.2.48)$$

hiermit ergibt sich

$$\begin{aligned} \hat{\beta}_1 &= (X_1^T X_1)^{-1} X_1^T \vec{Y} - (X_1^T X_1)^{-1} X_1^T X_2 \hat{\beta}_2 \\ \hat{\beta}_2 &= (X_2^T X_2)^{-1} X_2^T \vec{Y} - (X_2^T X_2)^{-1} X_2^T X_1 \hat{\beta}_1 \end{aligned} \quad (5.2.49)$$

Der BLE-Schätzer von  $\vec{\beta}_1$  für das reduzierte Modell

$$\hat{\beta}_{1,\omega} = (X_1^T X_1)^{-1} X_1^T \vec{Y} \quad (5.2.50)$$

stimmt mit dem BLE-Schätzer  $\hat{\beta}_1$  von  $\vec{\beta}_1$  für das vollständige Modell genau dann überein, wenn  $X_1^T X_2 = O$ , das heißt, wenn die Spalten von  $X_1$  zu den Spalten von  $X_2$  orthogonal sind. Genau dann stimmt auch der BLE-Schätzer von  $\vec{\beta}_2$

$$\hat{\beta}_{2,\tilde{\omega}} = (X_2^T X_2)^{-1} X_2^T \vec{Y} \quad (5.2.51)$$

für das reduzierte Modell

$$\vec{Y} = X_2 \vec{\beta}_2 + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \text{rg}(X_2) = k - p < k, \quad \vec{\beta}_2 \in \mathbb{R}^{k-p} \quad (5.2.52)$$

mit dem BLE-Schätzer  $\hat{\beta}_2$  von  $\vec{\beta}_2$  für das vollständige Modell überein und es gilt

$$(M^{22})^{-1} = X_2^T X_2. \quad (5.2.53)$$

Es sei  $SQ_R(\vec{\beta}_1, \vec{\beta}_2)$  die Modellquadratsumme des vollständigen Modells,  $SQ_R(\vec{\beta}_1)$  die Modellquadratsumme des durch Gleichung (5.2.46) und  $SQ_R(\vec{\beta}_2)$  die Modellquadratsumme des durch Gleichung (5.2.52) gegebenen reduzierten Modells sowie

$$SQ_R(\vec{\beta}_2 | \vec{\beta}_1) = SQ_R(\vec{\beta}_1, \vec{\beta}_2) - SQ_R(\vec{\beta}_1) \quad (5.2.54)$$

die Quadratsumme, um die die Modellquadratsumme des Modells, das nur  $\vec{\beta}_1$  enthält, zunimmt, wenn  $\vec{\beta}_2$  zusätzlich in das Modell aufgenommen wird. Dann ergibt sich

$$\begin{aligned} SQ_{R,\omega} &= SQ_R(\vec{\beta}_1) \\ SQ_{R,H} &= SQ_R(\vec{\beta}_2 | \vec{\beta}_1) \end{aligned} \quad (5.2.55)$$

Sind die Spalten von  $X_1$  orthogonal zu den Spalten von  $X_2$ , dann ergibt sich

$$\begin{aligned}
 SQ_R(\vec{\beta}_1, \vec{\beta}_2) &= \hat{Y}^T \hat{Y} = \hat{\beta}^T X^T X \hat{\beta} \\
 &= \hat{\beta}_1^T X_1^T X_1 \hat{\beta}_1 + \hat{\beta}_2^T X_2^T X_2 \hat{\beta}_2 + 2 \cdot \hat{\beta}_1^T X_1^T X_2 \hat{\beta}_2 \\
 &= \hat{\beta}_{1,\omega}^T X_1^T X_1 \hat{\beta}_{1,\omega} + \hat{\beta}_{2,\tilde{\omega}}^T X_2^T X_2 \hat{\beta}_{2,\tilde{\omega}} \\
 &= \hat{Y}_\omega^T \hat{Y}_\omega + \hat{Y}_{\tilde{\omega}}^T \hat{Y}_{\tilde{\omega}} \\
 &= SQ_R(\vec{\beta}_1) + SQ_R(\vec{\beta}_2)
 \end{aligned} \tag{5.2.56}$$

und somit

$$SQ_{R,H} = SQ_R(\vec{\beta}_2 | \vec{\beta}_1) = SQ_R(\vec{\beta}_2) . \tag{5.2.57}$$

### Beispiel 5.2.6. Vergleich zweier Regressionsgeraden

Beispiele für Arbeitshypothesen die den Anwendungsbereich kennzeichnen:

1. Im Stadtverkehr beeinflusst das Gewicht eines PKW' den Benzinverbrauch stärker als auf den Landstraßen.
2. Kinder, die bei Geburt relativ klein sind, wachsen in den ersten Lebensmonaten schneller als diejenigen, welche bei Geburt relativ groß sind.

Es ist die Auswirkung zweier unterschiedlicher Bedingungen, Behandlungen, Gruppenzugehörigkeiten - zwei Stufen eines qualitativen Einflussfaktors - auf die Abhängigkeit einer Größe von einer anderen - einer quantitativen Einflussgröße - zu untersuchen. Somit liegt ein Beispiel für eine Kovarianzanalyse vor.

Das durch die Gleichungen

$$\begin{aligned}
 Y_{1j} &= \beta_{10} + \beta_{11}x_{1j} + \epsilon_{1j} , \quad (j = 1, \dots, n_1) \\
 Y_{2j} &= \beta_{20} + \beta_{21}x_{2j} + \epsilon_{2j} , \quad (j = 1, \dots, n_2) \\
 \epsilon_{11}, \dots, \epsilon_{1n_1}, \epsilon_{21}, \dots, \epsilon_{2n_2} &\text{ unabhängig und jeweils } \mathfrak{N}(0, \sigma^2)\text{-verteilt}
 \end{aligned} \tag{5.2.58}$$

gegebene klassische lineare Modell lautet in Matrizen Schreibweise:

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I) \tag{5.2.59}$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\begin{aligned}
 \vec{Y} &= \begin{pmatrix} \vec{Y}_1 \\ \vec{Y}_2 \end{pmatrix}, \quad X = \begin{pmatrix} X_{11} & O \\ O & X_{22} \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \vec{\beta}_1 \\ \vec{\beta}_2 \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \end{pmatrix} \\
 \vec{Y}_i &= \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix}, \quad X_{ii} = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}, \quad \vec{\beta}_i = \begin{pmatrix} \beta_{i0} \\ \beta_{i1} \end{pmatrix}, \quad \vec{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}, \quad i = 1, 2.
 \end{aligned} \tag{5.2.60}$$

Da die Designmatrix  $X$  in die orthogonalen Teilmatrizen

$$X_1 = \begin{pmatrix} X_{11} \\ O \end{pmatrix}, \quad X_2 = \begin{pmatrix} O \\ X_{22} \end{pmatrix} \quad (5.2.61)$$

zerfällt, können die BLE-Schätzer von  $\vec{\beta}_1$  und  $\vec{\beta}_2$  nach Beispiel 5.2.5 als BLE-Schätzer der Teilmodelle

$$\vec{Y} = X_i \vec{\beta}_i + \vec{\epsilon} \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad i = 1, 2 \quad (5.2.62)$$

und somit der Einzelregressionen

$$\vec{Y}_i = X_{ii} \vec{\beta}_i + \vec{\epsilon}_i \quad \vec{\epsilon}_i \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad i = 1, 2 \quad (5.2.63)$$

bestimmt werden, da die Normalgleichungen der Teilmodelle und der Einzelregressionen übereinstimmen

$$X_i^T X_i \hat{\beta}_i = X_{ii}^T X_{ii} \hat{\beta}_i, \quad X_i^T \vec{Y} = X_{ii}^T \vec{Y}_i \quad (5.2.64)$$

Hieraus folgt

$$\begin{aligned} \hat{\beta}_{i0} &= \bar{Y}_i - \hat{\beta}_{i1} \bar{x}_i, \quad \hat{\beta}_{i1} = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(Y_{ij} - \bar{Y}_i)}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \\ \sigma_{\hat{\beta}_{i0}}^2 &= \frac{\sigma^2 \sum_{j=1}^{n_i} x_{ij}^2}{n \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}, \quad \sigma_{\hat{\beta}_{i1}}^2 = \frac{\sigma^2}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \\ \sigma_{\hat{\beta}_{i0}, \hat{\beta}_{i1}} &= -\frac{\sigma^2 \bar{x}_i}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}, \quad i = 1, 2 \\ \hat{\sigma}^2 &= \frac{1}{n_1 + n_2 - 4} \left( \sum_{j=1}^{n_1} (Y_{1j} - (\hat{\beta}_{10} + \hat{\beta}_{10} x_{1j}))^2 + \sum_{j=1}^{n_2} (Y_{2j} - (\hat{\beta}_{20} + \hat{\beta}_{20} x_{2j}))^2 \right) \\ &(\hat{\beta}_{10}, \hat{\beta}_{12})^T, (\hat{\beta}_{20}, \hat{\beta}_{21}), \hat{\sigma}^2 \text{ stochastisch unabhängig} \end{aligned} \quad (5.2.65)$$

#### a) Test auf Parallelität

Für die Hypothese der Parallelität der Regressionsgeraden

$$H : \beta_{11} = \beta_{21} \quad (5.2.66)$$

ergibt sich mit  $\psi(\vec{\beta}) = \beta_{11} - \beta_{21}$  die Darstellung

$$H : \psi(\vec{\beta}) = \beta_{11} - \beta_{21} = 0. \quad (5.2.67)$$

Aus Satz 5.2.1 über die t-Tests für eindimensionale schätzbare Funktionen ergibt sich mit

$$\sigma_{\hat{\beta}_{11} - \hat{\beta}_{21}}^2 = \sigma_{\hat{\beta}_{11}}^2 + \sigma_{\hat{\beta}_{21}}^2 = \left( \frac{1}{\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2} + \frac{1}{\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2} \right) \sigma^2 \quad (5.2.68)$$

der folgende t-Test. Die Hypothese  $H$  der Parallelität der Regressionsgeraden wird genau dann zum Niveau  $\alpha$  abgelehnt, wenn

$$|t| = \frac{|\hat{\beta}_{11} - \hat{\beta}_{21}|}{\hat{\sigma} \sqrt{\frac{1}{\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2} + \frac{1}{\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2}}} > t_{n_1 + n_2 - 4; 1 - \frac{\alpha}{2}} \quad (5.2.69)$$



**b) Test auf Gleichheit**

Für die Hypothese der Gleichheit der Regressionsgeraden

$$H : \beta_{10} = \beta_{20}, \beta_{11} = \beta_{21} \quad (5.2.70)$$

ergibt sich mit

$$C = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

$$\vec{\psi}(\vec{\beta}) = C\vec{\beta} \quad (5.2.71)$$

die Darstellung

$$H : \vec{\psi}(\vec{\beta}) = \begin{pmatrix} \beta_{10} - \beta_{20} \\ \beta_{11} - \beta_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (5.2.72)$$

Für die Quadratsumme der Hypothese ergibt sich

$$\begin{aligned} SQ_{R,H} &= SQ_R - SQ_{R,\omega} \\ &= SQ_R(\vec{\beta}_1) + SQ_R(\vec{\beta}_2) - SQ_{R,\omega} \\ &= \hat{\beta}_1^T X_1^T X_1 \hat{\beta}_1 + \hat{\beta}_2^T X_2^T X_2 \hat{\beta}_2 - SQ_{R,\omega} \\ &= \hat{\beta}_1^T X_{11}^T X_{11} \hat{\beta}_1 + \hat{\beta}_2^T X_{22}^T X_{22} \hat{\beta}_2 - SQ_{R,\omega} \\ &= SQ_{R,1} + SQ_{R,2} - SQ_{R,\omega} \end{aligned} \quad (5.2.73)$$

hierbei sind  $SQ_{R,1}$  und  $SQ_{R,2}$  die Modellquadratsummen der beiden durch Gleichung (5.2.63) gegebenen Einzelregressionen und  $SQ_{R,\omega}$  die Modellquadratsumme des reduzierten Modells

$$Y_{ij} = \gamma_0 + \gamma_1 x_{ij} + \epsilon_{ij}, \quad (j = 1, \dots, n_i, i = 1, 2) \quad (5.2.74)$$

$\epsilon_{11}, \dots, \epsilon_{1n_1}, \epsilon_{21}, \dots, \epsilon_{2n_2}$  unabhängig und jeweils  $\mathcal{N}(0, \sigma^2)$ -verteilt

Die BLE-Schätzer für das reduzierte Modell lauten

$$\hat{\gamma}_0 = \bar{Y}_{..} - \hat{\gamma}_1 \bar{x}_{..}, \quad \hat{\gamma}_1 = \frac{\sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})(Y_{ij} - \bar{Y}_{..})}{\sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2} \quad (5.2.75)$$

Für die Quadratsumme der Hypothese ergibt sich

$$\begin{aligned} SQ_{R,H} &= \hat{\beta}_1^T X_{11}^T \vec{Y}_1 + \hat{\beta}_2^T X_{22}^T \vec{Y}_2 - \hat{\gamma}^T [X_{11}^T X_{22}^T] \vec{Y} \\ &= \hat{\beta}_1^T X_{11}^T \vec{Y}_1 + \hat{\beta}_2^T X_{22}^T \vec{Y}_2 - \hat{\gamma}^T (X_{11}^T \vec{Y}_1 + X_{22}^T \vec{Y}_2) \\ &= (\hat{\beta}_{10} - \hat{\gamma}_0) Y_{1.} + (\hat{\beta}_{20} - \hat{\gamma}_0) Y_{2.} + (\hat{\beta}_{11} - \hat{\gamma}_1) \sum_{j=1}^{n_1} x_{1j} Y_{1j} + (\hat{\beta}_{21} - \hat{\gamma}_1) \sum_{j=1}^{n_2} x_{2j} Y_{2j} \\ &= (\hat{\beta}_\Omega - \hat{\beta}_\omega)^T X^T \vec{Y}, \quad \hat{\beta}_\Omega = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{20}, \hat{\beta}_{21})^T, \quad \hat{\beta}_\omega = (\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_0, \hat{\gamma}_1)^T \end{aligned} \quad (5.2.76)$$

Der F-Test auf Gleichheit der Regressionsgeraden führt zur Ablehnung der Hypothese  $H$  genau dann, wenn

$$F = \frac{SQ_{R,H}}{2\hat{\sigma}^2} > F_{2, n_1+n_2-4, 1-\alpha} \quad (5.2.77)$$



# Kapitel 6

## Varianzanalyse

### 6.1 Einfache Varianzanalyse

Wir betrachten im Folgenden die einfache Varianzanalyse, die im Beispiel 1.1.4 eingeführt wurde, und auch als  $k$ -Stichprobenproblem bezeichnet wird. Für die Varianzanalyse sind zwei unterschiedliche Darstellungen als klassisches lineares Modell gebräuchlich.

Das Modell mit einer Designmatrix mit vollem Spaltenrang ist durch die Gleichungen

$$\begin{aligned} Y_{ij} &= \mu_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1), \quad \sigma^2 > 0 \\ \epsilon_{ij} \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1) &\text{unabhängig}, \quad n = \sum_{i=1}^k n_i \end{aligned} \quad (6.1.1)$$

gegeben und lautet in Matrizenform mit  $\vec{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$  und  $\vec{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$  für  $i = 1, \dots, k$

$$\vec{Y} = \tilde{X}\vec{\gamma} + \vec{\epsilon} \quad \vec{\epsilon} \sim \mathcal{N}(\vec{0}, \sigma^2 I) \quad \sigma^2 > 0 \quad (6.1.2)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_k \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} \vec{1}_{n_1} & \vec{0}_{n_1} & \dots & \vec{0}_{n_1} \\ \vec{0}_{n_2} & \vec{1}_{n_2} & \dots & \vec{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{0}_{n_k} & \vec{0}_{n_k} & \dots & \vec{1}_{n_k} \end{pmatrix}, \quad \vec{\gamma} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \\ \vdots \\ \vec{\epsilon}_k \end{pmatrix} \quad (6.1.3)$$

Die Parameter  $\mu_i$  sind die Erwartungswerte bzw. (erwarteten) Mittelwerte der Stufen des zu untersuchenden Faktors, der den Stichproben zugrundeliegenden Grundgesamtheiten.

Das alternative lineare Modell verwendet eine Parametrisierung, bei der die Erwartungswerte in ein Gesamtmittel  $\mu$  und Effekte  $\alpha_i$  der Stufen des zu untersuchenden Faktors zerlegt werden. Die Modell-

gleichungen lauten dann

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathfrak{N}(0, \sigma^2) \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1), \quad \sigma^2 > 0$$

$$\epsilon_{ij} \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1) \text{ unabhängig}, \quad n = \sum_{i=1}^k n_i \quad (6.1.4)$$

und in Matrizenform

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I), \quad \sigma^2 > 0 \quad (6.1.5)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_k \end{pmatrix}, \quad X = \begin{pmatrix} \vec{1}_{n_1} & \vec{1}_{n_1} & \vec{0}_{n_1} & \dots & \vec{0}_{n_1} \\ \vec{1}_{n_2} & \vec{0}_{n_2} & \vec{1}_{n_2} & \dots & \vec{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{1}_{n_k} & \vec{0}_{n_k} & \vec{0}_{n_k} & \dots & \vec{1}_{n_k} \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \\ \vdots \\ \vec{\epsilon}_k \end{pmatrix} \quad (6.1.6)$$

Die Designmatrix  $X$  ist nicht von vollem Spaltenrang.

Im Beispiel 2.1.2 wurden die BLE-Schätzer für die Erwartungswerte  $\mu_i$ , die Parameter in der Modellgleichung (6.1.1), und deren Varianzen und Kovarianzen und in Beispiel 2.2.2 die Fehlerquadratsumme, auf der die Schätzung der Fehlervarianz basiert, bestimmt. Hieraus folgt zusammen mit den Ergebnissen aus Korollar 1 zu Satz 3.3.2 die Aussage des folgenden Satzes.

### Satz 6.1.1.

Für die BLE-Schätzer der Parameter des durch Gleichung (6.1.1) gegebenen Modells der einfachen Varianzanalyse gilt

1.  $\hat{\mu}_i = \bar{Y}_i = \frac{1}{n_i} Y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \sim \mathfrak{N}(\mu_i, \frac{\sigma^2}{n_i})$ ,  $i = 1, \dots, k$
2.  $\frac{n-k}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} SQ_E = \frac{1}{\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \frac{1}{\sigma^2} SQ_{in} \sim \chi_{n-k}^2$
3.  $\hat{\mu}_1, \dots, \hat{\mu}_k, \hat{\sigma}^2$  sind unabhängig.

Zusammen mit Satz 4.1.1 und Satz 5.2.1 ergibt sich das folgende Korollar sowie Satz 6.1.2.

### Korollar 1.

Für den BLE-Schätzer  $\hat{\psi}_{\vec{c}} = \sum_{i=1}^k c_i \bar{Y}_i$  von  $\vec{\psi}_{\vec{c}}(\vec{\gamma}) = \vec{c}^T \vec{\beta}_* = \sum_{i=1}^k c_i \mu_i$  mit  $\vec{c} = (c_1, \dots, c_k)^T \in \mathbb{R}^k$  gilt

1.  $\sum_{i=1}^k c_i \bar{Y}_i \sim \mathfrak{N}(\sum_{i=1}^k c_i \mu_i, \sigma^2 \sum_{i=1}^k \frac{c_i^2}{n_i})$
2.  $\sum_{i=1}^k c_i \bar{Y}_i$  und  $\hat{\sigma}^2 = \frac{1}{n-k} SQ_{in}$  sind unabhängig.
3.  $t = \frac{\sum_{i=1}^k c_i \bar{Y}_i}{\sqrt{\hat{\sigma}^2 \sum_{i=1}^k \frac{c_i^2}{n_i}}} \sim t_{n-k, \delta}$ ,  $\delta = \frac{\sum_{i=1}^k c_i \mu_i}{\sqrt{\sigma^2 \sum_{i=1}^k \frac{c_i^2}{n_i}}} \quad (\vec{c} \neq \vec{0})$
4.  $\frac{\sum_{i=1}^k c_i \bar{Y}_i - \sum_{i=1}^k c_i \mu_i}{\sqrt{\hat{\sigma}^2 \sum_{i=1}^k \frac{c_i^2}{n_i}}} \sim t_{n-k} \quad (\vec{c} \neq \vec{0})$

**Satz 6.1.2.**

**1. t-Test für eine Linearkombination der Erwartungswerte**

Die Hypothese

$$H : \sum_{i=1}^k c_i \mu_i = 0 \quad (\vec{c} \neq \vec{0}) \quad (6.1.7)$$

wird zugunsten der Alternative  $A : \sum_{i=1}^k c_i \mu_i \neq 0$  zum Niveau  $\alpha$  abgelehnt genau dann, wenn

$$\left| \frac{\sum_{i=1}^k c_i \bar{Y}_i}{\sqrt{\hat{\sigma}^2 \sum_{i=1}^k \frac{c_i^2}{n_i}}} \right| > t_{n-k; 1-\frac{\alpha}{2}} \quad (\vec{c} \neq \vec{0}) \quad (6.1.8)$$

**2. Konfidenzintervall für eine Linearkombination der Erwartungswerte**

Das Intervall

$$K(\vec{Y}) = \left[ \sum_{i=1}^k c_i \bar{Y}_i - t_{n-k; 1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2 \sum_{i=1}^k \frac{c_i^2}{n_i}}, \sum_{i=1}^k c_i \bar{Y}_i + t_{n-k; 1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2 \sum_{i=1}^k \frac{c_i^2}{n_i}} \right] \quad (6.1.9)$$

ist ein  $(1 - \alpha)$ -Konfidenzintervall für  $\sum_{i=1}^k c_i \mu_i$ .

**Beispiel 6.1.1.**

**1. Für die Hypothese**

$$H : \mu_1 = \mu_2 \quad (6.1.10)$$

ergibt sich mit  $\vec{c} = (1, -1, 0, \dots, 0)$  die Darstellung

$$H : \sum_{i=1}^k c_i \mu_i = \mu_1 - \mu_2 = 0 \quad (6.1.11)$$

$H$  wird zugunsten der Alternative  $A : \mu_1 \neq \mu_2$  zum Niveau  $\alpha$  abgelehnt genau dann, wenn

$$\left| \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \right| > t_{n-k; 1-\frac{\alpha}{2}} \quad (\vec{c} \neq \vec{0}) \quad (6.1.12)$$

**2. Für die Hypothese**

$$H : \mu_1 = \frac{\mu_2 + \mu_3}{2} \quad (6.1.13)$$

ergibt sich mit  $\vec{c} = (1, -\frac{1}{2}, -\frac{1}{2}, 0, \dots, 0)$  die Darstellung

$$H : \sum_{i=1}^k c_i \mu_i = \mu_1 - \frac{\mu_2}{2} - \frac{\mu_3}{2} = 0 \quad (6.1.14)$$

$H$  wird zugunsten der Alternative  $A : \mu_1 \neq \frac{\mu_2 + \mu_3}{2}$  zum Niveau  $\alpha$  abgelehnt genau dann, wenn

$$\left| \frac{\bar{Y}_1 - \frac{1}{2}\bar{Y}_2 - \frac{1}{2}\bar{Y}_3}{\sqrt{\hat{\sigma}^2(\frac{1}{n_1} + \frac{1}{4n_2} + \frac{1}{4n_3})}} \right| > t_{n-k; 1-\frac{\alpha}{2}} \quad (\vec{c} \neq \vec{0}) \quad (6.1.15)$$

### Satz 6.1.3. $F$ -Test auf Gleichheit der Effekte

Die Hypothese der Gleichheit der Erwartungswerte

$$H : \mu_1 = \mu_2 = \dots = \mu_k \quad (6.1.16)$$

in der Parametrisierung des linearen Modells ohne Rangdefekt (Gl. (6.1.1)) entspricht der Hypothese auf Gleichheit der Effekte

$$H : \alpha_1 = \alpha_2 = \dots = \alpha_k \quad (6.1.17)$$

in der Parametrisierung des linearen Modells mit Rangdefekt (Gl. (6.1.4)).

Das reduzierte Modell unter  $H$  ist

$$\begin{aligned} Y_{ij} &= \mu + \epsilon_{ij} \quad (j = 1, \dots, n_i > 0, i = 1, \dots, k > 1) \\ \vec{Y} &= \vec{1}_n \mu + \vec{\epsilon}, \quad \vec{\epsilon} \sim \mathcal{N}(\vec{0}, \sigma^2 I), \quad \mu \in \mathbb{R} \end{aligned} \quad (6.1.18)$$

somit ergibt sich mit Beispiel 5.2.4 der Test für  $H$  als  $F$ -Test auf Signifikanz des Modells mit der Testgröße (Gl. (5.2.26))

$$F = \frac{\frac{1}{k-1}(S_\omega - S_\Omega)}{\frac{1}{n-k}S_\Omega} = \frac{SQ_{Rc}}{(k-1)\hat{\sigma}^2}. \quad (6.1.19)$$

Mit der in Beispiel 2.2.2 durchgeführten Streuungszerlegung für die einfache Varianzanalyse ergibt sich

$$F = \frac{\frac{1}{k-1}(SQ_{Rc})}{\frac{1}{n-k}SQ_E} = \frac{\frac{1}{k-1}SQ_{zw}}{\frac{1}{n-k}SQ_{in}} = \frac{\frac{1}{k-1} \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}. \quad (6.1.20)$$

Die Hypothese  $H$  auf Gleichheit der Effekte der Stufen des untersuchten Faktors bzw. der Erwartungswerte der Stichproben wird zum Niveau  $\alpha$  abgelehnt genau dann, wenn  $F > F_{k-1, n-k; 1-\alpha}$ .

Die Tabelle der Varianzanalyse für die einfache Varianzanalyse hat die folgende Form:

Streuung	FG	SQ	MQ	F-Wert
zwischen den Faktorstufen	$k-1$	$SQ_{zw} = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2$	$MQ_{zw} = \frac{1}{k-1} SQ_{zw}$	$\frac{MQ_{zw}}{MQ_{in}}$
innerhalb der Faktorstufen	$n-k$	$SQ_{in} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$	$MQ_{in} = \frac{1}{n-k} SQ_{in}$	
Korr. Total	$n-1$	$SQ_{cT} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$		

### Beispiel 6.1.2. Simultane Konfidenzintervalle für Kontraste

#### 1. Scheffé-Methode der multiplen Vergleiche für Kontraste

Die Menge der linearen Kontraste  $\mathcal{L}'_K = \{\psi_{\vec{c}}(\vec{\gamma}) = \sum_{i=1}^k c_i \mu_i \mid \vec{c} = (c_1, \dots, c_k)^T \in \mathcal{L}_K\}$  mit  $\mathcal{L}_K = \{\vec{c} \in \mathbb{R}^k \mid \vec{c}^T \vec{1}_k = \sum_{i=1}^k c_i = 0\}$  bildet einen  $(k-1)$ -dimensionalen linearen Vektorraum, da  $\mathcal{L}_K = \mathcal{L}(\vec{1}_k)^\perp$ . Da die  $k-1$  Differenzen  $\mu_1 - \mu_i$  ( $i = 2, \dots, k$ ) linear unabhängige Kontraste  $\psi_i(\vec{\gamma}) = \vec{c}_i^T \vec{\gamma}$  mit  $\vec{c}_2 = (1, -1, 0, \dots, 0)^T$ ,  $\vec{c}_3 = (1, 0, -1, 0, \dots, 0)^T, \dots, \vec{c}_k = (1, 0, \dots, 0, -1)^T$  sind, bilden sie eine Basis von  $\mathcal{L}'_K$ . Aus Satz 4.2.3 folgt mit Korollar 1 zu Satz 6.1.1, dass  $(K_{\psi_{\vec{c}}})_{\psi_{\vec{c}} \in \mathcal{L}'_K}$  mit

$$\begin{aligned} K_{\psi_{\vec{c}}}(\vec{Y}) &= \left\{ \psi \mid \hat{\psi}_{\vec{c}} - \sqrt{(k-1)F_{k-1, n-k; 1-\alpha}} \hat{\sigma}_{\hat{\psi}_{\vec{c}}} \leq \psi \leq \hat{\psi}_{\vec{c}} + \sqrt{(k-1)F_{k-1, n-k; 1-\alpha}} \hat{\sigma}_{\hat{\psi}_{\vec{c}}} \right\} \\ &= \left[ \sum_{i=1}^k c_i \bar{Y}_i - d \sqrt{\frac{SQ_{in}}{n-k} \sum_{i=1}^k \frac{c_i^2}{n_i}}, \sum_{i=1}^k c_i \bar{Y}_i + d \sqrt{\frac{SQ_{in}}{n-k} \sum_{i=1}^k \frac{c_i^2}{n_i}} \right] \end{aligned} \quad (6.1.21)$$

mit  $d = \sqrt{(k-1)F_{k-1, n-k; 1-\alpha}}$ , bzw

$$K_{\psi_{\vec{c}}}(\vec{Y}) = \left[ \sum_{i=1}^k c_i \bar{Y}_i - d \|\vec{c}\| \sqrt{\frac{SQ_{in}}{(n-k)l}}, \sum_{i=1}^k c_i \bar{Y}_i + d \|\vec{c}\| \sqrt{\frac{SQ_{in}}{(n-k)l}} \right] \quad (6.1.22)$$

falls  $n_1 = n_2 = \dots = n_k = l$ , eine Familie von simultanen Konfidenzintervallen zum Niveau  $(1-\alpha)$  ( $0 < \alpha < 1$ ) für die linearen Kontraste der Erwartungswerte  $\mu_1, \dots, \mu_k$  und somit auch der Effekte  $\alpha_1, \dots, \alpha_k$  (siehe Seite 39) ist.

Für die Mittelwertsvergleiche  $\mu_i - \mu_j$  bzw. Effektvergleiche  $\alpha_i - \alpha_j$  lauten die Scheffé-Intervalle

$$K_{\mu_i - \mu_j}(\vec{Y}) = \left[ \bar{Y}_i - \bar{Y}_j - d \sqrt{\frac{SQ_{in}}{n-k} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \bar{Y}_i - \bar{Y}_j + d \sqrt{\frac{SQ_{in}}{n-k} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right] \quad (6.1.23)$$

#### 2. Bonferroni-Methode für multiple Mittelwertsvergleiche

##### 1. Vergleich mit einer Kontrollgruppe:

Die Beobachtungseinheiten der ersten Stichprobe stellen eine nichtbehandelte Kontrollgruppe dar, mit der die Effekte der verschiedenen Behandlungen, die auf die Beobachtungseinheiten der anderen  $k-1$  Stichproben angewandt wurden, verglichen werden sollen.

Aus Satz 4.2.1 folgt mit Satz 6.1.2, dass

$$\begin{aligned} K_{\mu_1 - \mu_i}(\vec{Y}) &= [\hat{\psi}_i - t_{n-k; 1-\frac{\alpha}{2(k-1)}} \hat{\sigma}_{\hat{\psi}_i}, \hat{\psi}_i + t_{n-k; 1-\frac{\alpha}{2(k-1)}} \hat{\sigma}_{\hat{\psi}_i}] \\ &= \left[ \bar{Y}_1 - \bar{Y}_i - d' \sqrt{\frac{SQ_{in}}{n-k} \left( \frac{1}{n_1} + \frac{1}{n_i} \right)}, \bar{Y}_1 - \bar{Y}_i + d' \sqrt{\frac{SQ_{in}}{n-k} \left( \frac{1}{n_1} + \frac{1}{n_i} \right)} \right] \end{aligned} \quad (6.1.24)$$

mit  $d' = t_{n-k; 1-\frac{\alpha}{2(k-1)}}$  simultane  $(1-\alpha)$ -Konfidenzintervalle für die  $k-1$  schätzbaren linearen Kontraste  $\psi_2(\vec{\gamma}) = \mu_1 - \mu_2 = \alpha_1 - \alpha_2, \dots, \psi_k(\vec{\gamma}) = \mu_1 - \mu_k = \alpha_1 - \alpha_k$  sind.

2. Vergleich aller Mittelwerte:

Aus Satz 4.2.1 folgt mit Satz 6.1.2, dass

$$\begin{aligned} K_{\mu_i - \mu_j}(\vec{Y}) &= [\hat{\psi}_{ij} - t_{n-k; 1 - \frac{\alpha}{k(k-1)}} \hat{\sigma}_{\hat{\psi}_{ij}}, \hat{\psi}_{ij} + t_{n-k; 1 - \frac{\alpha}{k(k-1)}} \hat{\sigma}_{\hat{\psi}_{ij}}] \\ &= \left[ \bar{Y}_{i.} - \bar{Y}_{j.} - d' \sqrt{\frac{SQ_{in}}{n-k} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \bar{Y}_{i.} - \bar{Y}_{j.} + d' \sqrt{\frac{SQ_{in}}{n-k} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right] \end{aligned} \quad (6.1.25)$$

mit  $d' = t_{n-k; 1 - \frac{\alpha}{k(k-1)}}$  simultane  $(1 - \alpha)$ -Konfidenzintervalle für die  $\frac{k(k-1)}{2}$  schätzbaren linearen Kontraste  $\psi_{ij}(\vec{\gamma}) = \mu_i - \mu_j = \alpha_i - \alpha_j$  ( $1 \leq i < j \leq k$ ) sind.



## 6.2 Zweifache Varianzanalyse

Die zweifache Varianzanalyse ist ein Auswertungsverfahren für Versuche, in denen die Wirkung von zwei qualitativen Faktoren untersucht werden soll. Angenommen es sollen verschiedene Weizensorten und verschiedene Düngemittel in ihrer Wirkung auf den Ernteertrag geprüft werden, dann ist einer der zu untersuchenden Faktoren der Faktor Sorte (Faktor  $A$ ), der andere Faktor Düngemittel (Faktor  $B$ ). Von dem Faktor  $A$  mögen  $I$ , von dem Faktor  $B$  mögen  $J$  Stufen in den Versuch einbezogen worden sein. Bei Versuchen mit zwei Faktoren wird das Versuchsmaterial in zwei Richtungen klassifiziert; es liegt eine sogenannte Zweifachklassifikation vor.

Bezeichnet man die Kombinationen der Faktorstufen (auch Klassen oder Zellen genannt) mit  $(i, j)$  ( $i = 1, \dots, I; j = 1, \dots, J$ ), wobei  $n_{ij}$  die Anzahl der Beobachtungswerte der Klasse  $(i, j)$  bezeichnet, so kann man verschiedene Arten der Zweifachklassifikation wie folgt charakterisieren:

### 1. Vollständige Kreuzklassifikation:

$$n_{ij} \geq 1 \text{ für alle Klassen } (i, j)$$

Jede Stufe des Faktors  $A$  tritt in den Beobachtungswerten mit jeder Stufe des Faktors  $B$  auf. Alle  $I \cdot J$  Klassen (Zellen) sind besetzt. Sind alle  $n_{ij} = K \geq 1$ , so spricht man von einer Kreuzklassifikation mit gleicher Klassenbesetzung bzw. von einem ausgewogenen oder **balancierten Versuchsplan**.

### 2. Unvollständige Kreuzklassifikation:

$$\exists i, j_1 \neq j_2 : n_{ij_1}, n_{ij_2} > 0, \quad \exists j, i_1 \neq i_2 : n_{i_1j}, n_{i_2j} > 0, \quad \exists i, j : n_{ij} = 0$$

Mindestens eine Stufe jedes Faktors tritt in den Beobachtungswerten mit zwei Stufen des anderen Faktors auf, es liegt jedoch keine vollständige Kreuzklassifikation vor.

### 3. Hierarchische Klassifikation

$$n_{kj} > 0 \Rightarrow \forall i \neq k : n_{ij} = 0$$

In den Beobachtungswerten tritt jede Stufe des Faktors  $B$  nur mit einer Stufe des Faktors  $A$  auf. Der Faktor  $B$  ist dem Faktor  $A$  untergeordnet.

Wir betrachten im Folgenden die vollständige Kreuzklassifikation.

Bezeichnet  $Y_{ijk}$ ,  $k = 1, \dots, n_{ij}$  die  $k$ -te Beobachtung der Klasse  $(i, j)$ , und wird vorausgesetzt, dass die Beobachtungen unabhängig und normalverteilt mit derselben Varianz sind, dann können die Beobachtungen durch das folgende klassische lineare Modell beschrieben werden:

$$\begin{aligned} Y_{ijk} &= \mu_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \mathfrak{N}(0, \sigma^2), \quad \sigma^2 > 0 \\ (k &= 1, \dots, n_{ij} > 0, i = 1, \dots, I > 1, j = 1, \dots, J > 1) \\ \epsilon_{ijk} &\text{ unabhängig, } n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \end{aligned} \tag{6.2.1}$$

dieses lautet in Matrizenform mit  $\vec{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijn_{ij}})^T$  und  $\vec{\epsilon}_{ij} = (\epsilon_{ij1}, \dots, \epsilon_{ijn_{ij}})^T$  für  $i = 1, \dots, I$ ,  $j = 1, \dots, J$

$$\vec{Y} = \tilde{X}\vec{\gamma} + \vec{\epsilon} \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I) \quad \sigma^2 > 0 \quad (6.2.2)$$

wobei die Matrix und die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} \vec{Y}_{11} \\ \vec{Y}_{12} \\ \vdots \\ \vec{Y}_{1J} \\ \vec{Y}_{21} \\ \vdots \\ \vec{Y}_{IJ} \end{pmatrix}, \quad \vec{\gamma} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1J} \\ \mu_{21} \\ \vdots \\ \mu_{IJ} \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \vec{\epsilon}_{11} \\ \vec{\epsilon}_{12} \\ \vdots \\ \vec{\epsilon}_{1J} \\ \vec{\epsilon}_{21} \\ \vdots \\ \vec{\epsilon}_{IJ} \end{pmatrix} \quad (6.2.3)$$

$$\tilde{X} = \begin{pmatrix} \vec{1}_{n_{11}} & \vec{0}_{n_{11}} & \dots & \vec{0}_{n_{11}} & \vec{0}_{n_{11}} & \dots & \vec{0}_{n_{11}} \\ \vec{0}_{n_{12}} & \vec{1}_{n_{12}} & \dots & \vec{0}_{n_{12}} & \vec{0}_{n_{12}} & \dots & \vec{0}_{n_{12}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vec{0}_{n_{1J}} & \vec{0}_{n_{1J}} & \dots & \vec{1}_{n_{1J}} & \vec{0}_{n_{1J}} & \dots & \vec{0}_{n_{1J}} \\ \vec{0}_{n_{21}} & \vec{0}_{n_{21}} & \dots & \vec{0}_{n_{21}} & \vec{1}_{n_{21}} & \dots & \vec{0}_{n_{21}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vec{0}_{n_{IJ}} & \vec{0}_{n_{IJ}} & \dots & \vec{0}_{n_{IJ}} & \vec{0}_{n_{IJ}} & \dots & \vec{1}_{n_{IJ}} \end{pmatrix}$$

Die Designmatrix besitzt vollen Spaltenrang,  $rg(\tilde{X}) = IJ$ . Die Beobachtungen einer Klasse können als Beobachtungen einer Stichprobe aufgefasst werden. Die Parameter  $\mu_{ij}$  sind die Erwartungswerte bzw. (erwarteten) Mittelwerte der Klassen der zu untersuchenden Faktoren, der den Stichproben zugrundeliegenden Grundgesamtheiten.

Offensichtlich ist die  $(n \times (IJ))$ -Designmatrix  $\tilde{X}$  von derselben Bauart wie bei der einfachen Varianzanalyse, auch Einfachklassifikation genannt (siehe Gleichung (6.1.3)). Somit können die Aussagen von Satz 6.1.1 auf die zweifache Varianzanalyse übertragen werden:

### Satz 6.2.1.

Für die BLE-Schätzer der Parameter des durch Gleichung (6.2.1) gegebenen Modells der zweifachen Varianzanalyse gilt

1.  $\hat{\mu}_{ij} = \bar{Y}_{ij.} = \frac{1}{n_{ij}} Y_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} \sim \mathfrak{N}(\mu_{ij}, \frac{\sigma^2}{n_{ij}})$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$
2.  $\frac{n-IJ}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} SQ_E = \frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 \sim \chi_{n-IJ}^2$ , für  $n > IJ$
3.  $\hat{\mu}_{11}, \dots, \hat{\mu}_{IJ}, \hat{\sigma}^2$  sind unabhängig.

Ziel der zweifachen Varianzanalyse ist es, die Einflüsse der Faktoren  $A$  und  $B$  voneinander zu isolieren. Hierzu bedient man sich spezieller schätzbarer Funktionen.

**Definition 6.2.1.**

Bei der Zweifachklassifikation heit

$$\mu^* = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \mu_{ij} \quad (6.2.4)$$

allgemeines Mittel oder Gesamtmittel,

$$\alpha_i^* = \bar{\mu}_{i.} - \mu^* = \frac{1}{J} \sum_{j=1}^J \mu_{ij} - \mu^* \quad (6.2.5)$$

Haupteffekt der  $i$ -ten Stufe des Faktors  $A$ ,

$$\beta_j^* = \bar{\mu}_{.j} - \mu^* = \frac{1}{I} \sum_{i=1}^I \mu_{ij} - \mu^* \quad (6.2.6)$$

Haupteffekt der  $j$ -ten Stufe des Faktors  $B$ ,

$$\alpha_{i|j}^* = \mu_{ij} - \bar{\mu}_{.j} \quad (6.2.7)$$

Effekt der  $i$ -ten Stufe des Faktors  $A$  unter der Bedingung, dass Faktor  $B$  in der  $j$ -ten Stufe auftritt,

$$\beta_{j|i}^* = \mu_{ij} - \bar{\mu}_{i.} \quad (6.2.8)$$

Effekt der  $j$ -ten Stufe des Faktors  $B$  unter der Bedingung, dass Faktor  $A$  in der  $i$ -ten Stufe auftritt,

$$\gamma_{ij}^* = \alpha_{i|j}^* - \alpha_i^* = \beta_{j|i}^* - \beta_j^* = \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \mu^* \quad (6.2.9)$$

Wechselwirkung der  $i$ -ten Stufe des Faktors  $A$  mit der  $j$ -ten Stufe des Faktors  $B$ .

**Bemerkung 1.**

Aus der Definition der Haupteffekte und Wechselwirkungen folgt

1.  $\mu_{ij} = \mu^* + \alpha_i^* + \beta_j^* + \gamma_{ij}^*$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$
2.  $\sum_{i=1}^I \alpha_i^* = 0$ ,  $\sum_{j=1}^J \beta_j^* = 0$ ,  $\sum_{i=1}^I \gamma_{ij}^* = 0 \forall j$ ,  $\sum_{j=1}^J \gamma_{ij}^* = 0 \forall i$
3. Die BLE-Schtzer bezglich des durch Gleichung (6.2.1) gegebenen Modells der zweifachen Varianzanalyse ergeben sich zu

$$\begin{aligned} \hat{\mu}^* &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \bar{Y}_{ij.}, \quad \hat{\alpha}_i^* = \frac{1}{J} \sum_{j=1}^J \bar{Y}_{ij.} - \hat{\mu}^*, \quad \hat{\beta}_j^* = \frac{1}{I} \sum_{i=1}^I \bar{Y}_{ij.} - \hat{\mu}^* \\ \hat{\gamma}_{ij}^* &= \bar{Y}_{ij.} - \frac{1}{J} \sum_{j=1}^J \bar{Y}_{ij.} - \frac{1}{I} \sum_{i=1}^I \bar{Y}_{ij.} + \hat{\mu}^* \end{aligned} \quad (6.2.10)$$

4. Für einen balancierten Versuchsplan mit  $K > 0$  Beobachtungen pro Klasse lauten die BLE-Schätzer

$$\begin{aligned}\hat{\mu}^* &= \bar{Y}_{...} = \frac{1}{IJK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K Y_{ijk} \\ \hat{\alpha}_i^* &= \bar{Y}_{i..} - \bar{Y}_{...} = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K Y_{ijk} - \bar{Y}_{...} \\ \hat{\beta}_j^* &= \bar{Y}_{.j.} - \bar{Y}_{...} = \frac{1}{IK} \sum_{i=1}^I \sum_{k=1}^K Y_{ijk} - \bar{Y}_{...} \\ \hat{\gamma}_{ij}^* &= \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\end{aligned}\tag{6.2.11}$$

Die Beziehung  $\mu_{ij} = \mu^* + \alpha_i^* + \beta_j^* + \gamma_{ij}^*$  legt die folgende alternative Parametrisierung des linearen Modells nahe

$$\begin{aligned}Y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \mathfrak{N}(0, \sigma^2), \quad \sigma^2 > 0 \\ (k &= 1, \dots, n_{ij} > 0, i = 1, \dots, I > 1, j = 1, \dots, J > 1)\end{aligned}\tag{6.2.12}$$

$$\epsilon_{ijk} \text{ unabhängig}, \quad n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$$

dieses lautet in Matrizenform mit  $\vec{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijn_{ij}})^T$  und  $\vec{\epsilon}_{ij} = (\epsilon_{ij1}, \dots, \epsilon_{ijn_{ij}})^T$  für  $i = 1, \dots, I, j = 1, \dots, J$

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I) \quad \sigma^2 > 0\tag{6.2.13}$$

wobei die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} \vec{Y}_{11} \\ \vec{Y}_{12} \\ \vdots \\ \vec{Y}_{1J} \\ \vec{Y}_{21} \\ \vdots \\ \vec{Y}_{IJ} \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_I \\ \beta_1 \\ \vdots \\ \beta_J \\ \gamma_{11} \\ \vdots \\ \gamma_{IJ} \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \vec{\epsilon}_{11} \\ \vec{\epsilon}_{12} \\ \vdots \\ \vec{\epsilon}_{1J} \\ \vec{\epsilon}_{21} \\ \vdots \\ \vec{\epsilon}_{IJ} \end{pmatrix}\tag{6.2.14}$$

Die Designmatrix ist eine  $(n \times (1 + I + J + IJ))$ -Matrix. Da es aufgrund von Bemerkung 1 zu Definition 6.2.1 zu jedem Vektor  $\vec{\gamma} = (\mu_{11}, \dots, \mu_{IJ})^T$  einen Vektor  $\vec{\beta} = (\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \gamma_{11}, \dots, \gamma_{IJ})^T$  und mit  $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$  auch zu jedem  $\vec{\beta}$  einen Vektor  $\vec{\gamma}$  mit  $E[\vec{Y}] = \tilde{X}\vec{\gamma} = X\vec{\beta}$

gibt, stimmen der Spaltenraum von  $\tilde{X}$  und  $X$  überein, und es ist  $rg(X) = IJ$ .

Zwischen den Parametern der beiden alternativen Darstellungen des linearen Modells besteht die Beziehung

$$E[Y_{ijk}] = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad (6.2.15)$$

Diese Darstellung und damit auch die alternative Parametrisierung ist eindeutig, wenn bei der alternativen Parametrisierung das Erfülltsein der sogenannten Reparametrisierungsbedingungen

$$\sum_{i=1}^I \alpha_i = 0, \quad \sum_{j=1}^J \beta_j = 0, \quad \sum_{i=1}^I \gamma_{ij} = 0 \quad \forall j, \quad \sum_{j=1}^J \gamma_{ij} = 0 \quad \forall i \quad (6.2.16)$$

verlangt wird. Die Eindeutigkeit kann gezeigt werden, indem in den in Definition 6.2.1 definierten Größen  $\mu_{ij}$  durch  $\mu + \alpha_i + \beta_j + \gamma_{ij}$  ersetzt wird. So ergibt sich z.B.:

$$\begin{aligned} \mu^* &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \mu_{ij} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J (\mu + \alpha_i + \beta_j + \gamma_{ij}) \\ &= \mu + \frac{1}{I} \sum_{i=1}^I \alpha_i + \frac{1}{J} \sum_{j=1}^J \beta_j + \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \gamma_{ij} = \mu \end{aligned} \quad (6.2.17)$$

Die BLE-Schätzer der Parameter, die die in Gleichung (6.2.16) gegebenen Reparametrisierungsbedingungen erfüllen, bzw. die BLE-Schätzer der schätzbaren Funktionen  $\mu^*$ ,  $\alpha_i^*$ ,  $\beta_j^*$  und  $\gamma_{ij}^*$  sind die eindeutig bestimmten Parameterwerte, die die Summe der Quadrate der Abweichungen

$$S(\vec{Y}, \vec{\beta}) := \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^{n_{ij}} (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2, \quad (6.2.18)$$

minimieren und die Reparametrisierungsbedingungen erfüllen.

Für eine balancierte vollständige Kreuzklassifikation mit  $K > 0$  Beobachtungen pro Klasse, lassen sich diese BLE-Schätzer direkt aus der folgenden Zerlegung von  $S(\vec{Y}, \vec{\beta})$  ablesen, die sich unter Verwendung der Identität

$$\bar{Y}_{ij.} = \bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$$

ergibt. Dieses gilt, wie sich zeigen wird, auch für die BLE-Schätzer bestimmter reduzierter Modelle. Es gilt

$$\begin{aligned} S(\vec{Y}, \vec{\beta}) &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K ((Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{...} - \mu) + ((\bar{Y}_{i..} - \bar{Y}_{...}) - \alpha_i) \\ &\quad + ((\bar{Y}_{.j.} - \bar{Y}_{...}) - \beta_j) + ((\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) - \gamma_{ij}))^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{ij.})^2 + IJK(\bar{Y}_{...} - \mu)^2 + JK \sum_{i=1}^I ((\bar{Y}_{i..} - \bar{Y}_{...}) - \alpha_i)^2 \\ &\quad + IK \sum_{j=1}^J ((\bar{Y}_{.j.} - \bar{Y}_{...}) - \beta_j)^2 + K \sum_{i=1}^I \sum_{j=1}^J ((\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) - \gamma_{ij})^2, \end{aligned} \quad (6.2.19)$$

da die Summe aller gemischten Produkte unter anderem aufgrund der Reparametrisierungsbedingungen verschwinden.

Das durch die Gleichungen (6.2.1) bzw. (6.2.12) gegebenen Modell der zweifachen Varianzanalyse wird als Modell mit Wechselwirkungen bezeichnet.

**Definition 6.2.2.**

Die Effekte der beiden Faktoren  $A$  und  $B$  einer Zweifachklassifikation heißen *additiv genau dann, wenn*

$$\gamma_{ij}^* = 0, \quad \forall(i, j). \quad (6.2.20)$$

Geht man davon aus, dass die Effekte additiv sind, dann liegt ein Modell ohne Wechselwirkungen vor. Das zugehörige klassische lineare Modell hat die folgende Form:

$$\begin{aligned} Y_{ijk} &= \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \mathfrak{N}(0, \sigma^2), \quad \sigma^2 > 0 \\ (k &= 1, \dots, n_{ij} > 0, i = 1, \dots, I > 1, j = 1, \dots, J > 1) \\ \epsilon_{ijk} &\text{ unabhängig, } \quad n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \end{aligned} \quad (6.2.21)$$

dieses lautet in Matrizenform mit  $\vec{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijn_{ij}})^T$  und  $\vec{\epsilon}_{ij} = (\epsilon_{ij1}, \dots, \epsilon_{ijn_{ij}})^T$  für  $i = 1, \dots, I, j = 1, \dots, J$

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon} \quad \vec{\epsilon} \sim \mathfrak{N}(\vec{0}, \sigma^2 I) \quad \sigma^2 > 0 \quad (6.2.22)$$

wobei die Vektoren gegeben sind durch

$$\vec{Y} = \begin{pmatrix} \vec{Y}_{11} \\ \vec{Y}_{12} \\ \vdots \\ \vec{Y}_{1J} \\ \vec{Y}_{21} \\ \vdots \\ \vec{Y}_{IJ} \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_I \\ \beta_1 \\ \vdots \\ \beta_J \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \vec{\epsilon}_{11} \\ \vec{\epsilon}_{12} \\ \vdots \\ \vec{\epsilon}_{1J} \\ \vec{\epsilon}_{21} \\ \vdots \\ \vec{\epsilon}_{IJ} \end{pmatrix} \quad (6.2.23)$$

**Satz 6.2.2.**

Für das durch Gleichung (6.2.21) gegebene Modell einer zweifachen Varianzanalyse ohne Wechselwirkungen sei  $\mu_{ij} = E[Y_{ijk}]$  und das Gesamtmittel  $\mu^*$  und die Effekte  $\alpha_i^*$ ,  $\beta_j^*$  und  $\gamma_{ij}^*$  wie in Definition 6.2.1 definiert. Dann gilt

1.  $\mu_{ij} = \mu^* + \alpha_i^* + \beta_j^* = \mu + \alpha_i + \beta_j, \quad \gamma_{ij}^* = 0, \quad i = 1, \dots, I, j = 1, \dots, J$
2.  $\sum_{i=1}^I \alpha_i^* = 0, \quad \sum_{j=1}^J \beta_j^* = 0$

3. Die Darstellung  $\mu_{ij} = \mu + \alpha_i + \beta_j$  ist unter den Reparametrisierungsbedingungen

$$\sum_{i=1}^I \alpha_i = 0, \quad \sum_{j=1}^J \beta_j = 0 \quad (6.2.24)$$

eindeutig. Somit ist der Rang der  $(n \times (I + J + 1))$ -Designmatrix  $X$  gleich  $r = I + J - 1$ .

4. Die Parameterfunktionen  $\mu^*$ ,  $\alpha_i^*$  und  $\beta_j^*$  sind als Linearkombinationen der Erwartungswerte  $\mu_{ij}$  schätzbare Funktionen. Da

$$\begin{aligned} \sum_{i=1}^I c_i \alpha_i^* &= \sum_{i=1}^I c_i \left( \frac{1}{J} \sum_{j=1}^J (\mu + \alpha_i + \beta_j) - \frac{1}{IJ} \sum_{i'=1}^I \sum_{j=1}^J (\mu + \alpha_{i'} + \beta_j) \right) \\ &= \sum_{i=1}^I c_i \alpha_i \quad \text{für} \quad \sum_{i=1}^I c_i = 0 \\ \sum_{j=1}^J d_j \beta_j^* &= \sum_{j=1}^J d_j \beta_j \quad \text{für} \quad \sum_{j=1}^J d_j = 0 \end{aligned} \quad (6.2.25)$$

sind auch die Kontraste  $\sum_{i=1}^I c_i \alpha_i$  der Haupteffekte der Stufen des Faktors  $A$  und  $\sum_{j=1}^J d_j \beta_j$  der Haupteffekte der Stufen des Faktors  $B$  schätzbar.

5. Für einen balancierten Versuchsplan mit  $K > 0$  Beobachtungen pro Klasse lauten die BLE-Schätzer für die eindeutig bestimmten Parameterwerte, die den Reparametrisierungsbedingungen genügen

$$\begin{aligned} \hat{\mu}^* &= \bar{Y}_{...} = \frac{1}{IJK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K Y_{ijk} \\ \hat{\alpha}_i^* &= \bar{Y}_{i..} - \bar{Y}_{...} = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K Y_{ijk} - \bar{Y}_{...} \\ \hat{\beta}_j^* &= \bar{Y}_{.j.} - \bar{Y}_{...} = \frac{1}{IK} \sum_{i=1}^I \sum_{k=1}^K Y_{ijk} - \bar{Y}_{...} \end{aligned} \quad (6.2.26)$$

Hiermit ergibt sich die Fehlerquadratsumme zu

$$\begin{aligned} S_{\Omega} &= SQ_E = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \hat{\mu}^* - \hat{\alpha}_i^* - \hat{\beta}_j^*)^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \end{aligned} \quad (6.2.27)$$

### Satz 6.2.3. F-Tests für die balancierte zweifache Varianzanalyse ohne Wechselwirkungen

Es liege das durch Gleichung (6.2.21) gegebene Modell einer zweifachen Varianzanalyse ohne Wechselwirkungen mit  $K > 0$  Beobachtungen pro Klasse vor.

## 1. Die Hypothese

$$H_A : \alpha_1 = \alpha_2 = \dots = \alpha_I \quad (6.2.28)$$

kann auch als  $H_A : \alpha_1^* = \alpha_2^* = \dots = \alpha_I^* = 0$  sowie  $H_A : \alpha_1 - \alpha_i = 0$  ( $i = 2, \dots, I$ ) formuliert werden. Somit ist  $H_A$  eine testbare Hypothese mit  $q = I - 1$ , da die schätzbaren Funktionen  $\alpha_1 - \alpha_i$  ( $i = 2, \dots, I$ ) linear unabhängig sind. Das reduzierte Modell unter  $H_A$

$$Y_{ijk} = \mu + \beta_j + \epsilon_{ijk} \quad (6.2.29)$$

ist vom Typ der einfachen Varianzanalyse. Somit ist der BLE Schätzer von  $\mu_j = \mu + \beta_j$  unter  $H_A$  gleich  $\hat{\mu}_{j;\omega_A} = \bar{Y}_{.j} = \hat{\mu}^* + \hat{\beta}_j^*$ . Die Quadratsumme der Hypothese  $SQ_{R,H_A} = S_{\omega_A} - S_{\Omega}$  ergibt sich zu

$$\begin{aligned} SQ_{R,H_A} &= \|\hat{Y}_{\Omega} - \hat{Y}_{\omega_A}\|^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\hat{\mu}^* + \hat{\alpha}_i^* + \hat{\beta}_j^* - (\hat{\mu}^* + \hat{\beta}_j^*))^2 \\ &= JK \sum_{i=1}^I (\bar{Y}_{i..} - \bar{Y}_{...})^2 \end{aligned} \quad (6.2.30)$$

Der F-Test für  $H_A$  zum Niveau  $\alpha$  führt zur Ablehnung von  $H_A$  genau dann, wenn

$$F_A = \frac{\frac{1}{I-1} JK \sum_{i=1}^I (\bar{Y}_{i..} - \bar{Y}_{...})^2}{\frac{1}{IJK-I-J+1} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2} > F_{I-1, IJK-I-J+1; 1-\alpha} \quad (6.2.31)$$

## 2. Der F-Test für die Hypothese

$$H_B : \beta_1 = \beta_2 = \dots = \beta_J \quad (6.2.32)$$

zum Niveau  $\alpha$  führt zur Ablehnung von  $H_B$  genau dann, wenn

$$F_B = \frac{\frac{1}{J-1} IK \sum_{j=1}^J (\bar{Y}_{.j} - \bar{Y}_{...})^2}{\frac{1}{IJK-I-J+1} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2} > F_{J-1, IJK-I-J+1; 1-\alpha} \quad (6.2.33)$$

**Satz 6.2.4. F-Test auf Wechselwirkungen für die balancierte zweifache Varianzanalyse**

Es liege das durch Gleichung (6.2.12) gegebene Modell einer zweifachen Varianzanalyse mit Wechselwirkungen mit  $K > 0$  Beobachtungen pro Klasse vor. Das Gesamtmittel  $\mu^*$  und die Effekte  $\alpha_i^*$ ,  $\beta_j^*$  und  $\gamma_{ij}^*$  seien wie in Definition 6.2.1 definiert. Das reduzierte Modell unter der Hypothese

$$H_C : \gamma_{ij}^* = 0 \quad i = 1, \dots, I; j = 1, \dots, J \quad (6.2.34)$$



ist das durch Gleichung (6.2.21) gegebene Modell der zweifachen Varianzanalyse ohne Wechselwirkungen. Somit ergibt sich die Quadratsumme der Hypothese  $SQ_{R,H_C} = S_{\omega_C} - S_{\Omega}$  zu

$$\begin{aligned} SQ_{R,H_C} &= \|\hat{Y}_{\Omega} - \hat{Y}_{\omega_C}\|^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\hat{\mu}^* + \hat{\alpha}_i^* + \hat{\beta}_j^* + \hat{\gamma}_{ij}^* - (\hat{\mu}^* + \hat{\alpha}_i^* + \hat{\beta}_j^*))^2 \\ &= K \sum_{i=1}^I \sum_{j=1}^J (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \end{aligned} \quad (6.2.35)$$

Mit  $r - q = I + J - 1$  und  $r = IJ$  ergibt sich  $q = IJ - I - J + 1 = (I - 1)(J - 1)$ . Der  $F$ -Test für  $H_C$  zum Niveau  $\alpha$  führt zur Ablehnung von  $H_C$  genau dann, wenn

$$F_C = \frac{\frac{1}{(I-1)(J-1)} K \sum_{i=1}^I \sum_{j=1}^J (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2}{\frac{1}{IJ(K-1)} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{ij.})^2} > F_{(I-1)(J-1), IJ(K-1); 1-\alpha} \quad (6.2.36)$$

**Satz 6.2.5. F-Test auf Haupteffekte für die balancierte zweifache Varianzanalyse mit Wechselwirkungen**

Es liege das durch Gleichung (6.2.12) gegebene Modell einer zweifachen Varianzanalyse mit Wechselwirkungen mit  $K > 0$  Beobachtungen pro Klasse vor. Das Gesamtmittel  $\mu^*$  und die Effekte  $\alpha_i^*$ ,  $\beta_j^*$  und  $\gamma_{ij}^*$  seien wie in Definition 6.2.1 definiert.

1. Die Hypothese

$$H_A : \alpha_1^* = \alpha_2^* = \dots = \alpha_I^* = 0 \quad (6.2.37)$$

kann auch als  $H_A : \alpha_1^* - \alpha_i^* = 0$  ( $i = 2, \dots, I$ ) formuliert werden. Da  $\alpha_1^* - \alpha_i^* = \bar{\mu}_{1.} - \bar{\mu}_{i.}$  ( $i = 2, \dots, I$ ), sind die schätzbaren Funktionen  $\alpha_1^* - \alpha_i^*$  ( $i = 2, \dots, I$ ) linear unabhängig, somit ist  $H_A$  eine testbare Hypothese mit  $q = I - 1$ . Durch Minimierung der Summe der Quadrate der Abweichungen in Gleichung (6.2.19) bezüglich der Parameter des durch die Hypothese  $H_A$  gegebenen reduzierten Modelles, das heißt, Minimierung bezüglich  $\mu$ ,  $\beta_j$  und  $\gamma_{ij}$  unter den Bedingungen

$$\alpha_i = 0 \quad \forall i, \quad \sum_{j=1}^J \beta_j = 0, \quad \sum_{i=1}^I \gamma_{ij} = 0 \quad \forall j, \quad \sum_{j=1}^J \gamma_{ij} = 0 \quad \forall i \quad (6.2.38)$$

ergibt sich für das reduzierte Modell die Residuenquadratsumme

$$S_{\omega_A} = S_{\Omega} + JK \sum_{i=1}^I (\hat{\alpha}_i^*)^2 \quad (6.2.39)$$

Die Quadratsumme der Hypothese ist dann

$$SQ_{R,H_A} = S_{\omega_A} - S_{\Omega} = JK \sum_{i=1}^I (\bar{Y}_{i..} - \bar{Y}_{...})^2 \quad (6.2.40)$$

Der  $F$ -Test für  $H_A$  zum Niveau  $\alpha$  führt zur Ablehnung von  $H_A$  genau dann, wenn

$$F_A = \frac{\frac{1}{I-1} JK \sum_{i=1}^I (\bar{Y}_{i..} - \bar{Y}_{...})^2}{\frac{1}{IJ(K-1)} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{ij.})^2} > F_{I-1, IJ(K-1); 1-\alpha} \quad (6.2.41)$$

2. Der  $F$ -Test für die Hypothese

$$H_B : \beta_1^* = \beta_2^* = \dots = \beta_J^* = 0 \quad (6.2.42)$$

zum Niveau  $\alpha$  führt zur Ablehnung von  $H_B$  genau dann, wenn

$$F_B = \frac{\frac{1}{J-1} IK \sum_{j=1}^J (\bar{Y}_{.j.} - \bar{Y}_{...})^2}{\frac{1}{IJ(K-1)} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{ij.})^2} > F_{J-1, IJ(K-1); 1-\alpha} \quad (6.2.43)$$

**Bemerkung 1.**

Aus Gleichung (6.2.19) folgt mit  $\alpha_i = \beta_j = \gamma_{ij} = 0 \forall i, j$  durch Minimierung bezüglich  $\mu$

$$\begin{aligned} SQ_{cT} &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{...})^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{ij.})^2 + JK \sum_{i=1}^I (\hat{\alpha}_i^*)^2 + IK \sum_{j=1}^J (\hat{\beta}_j^*)^2 + K \sum_{i=1}^I \sum_{j=1}^J (\hat{\gamma}_{ij}^*)^2 \\ &= SQ_E + SQ_{R, H_A} + SQ_{R, H_B} + SQ_{R, H_C} \end{aligned} \quad (6.2.44)$$