

## 4 Einführung in die Theorie linearer Modelle

Häufig werden mehrere erklärende Variablen bzw. Regressoren  $X_1, \dots, X_p$  gleichzeitig betrachtet. Dafür benötigen wir eine Modellerweiterung.

### 4.1 Vorbemerkungen

#### Multivariate Zufallsvariablen

- Sei  $\mathbf{x}$  Vektor aus  $p$  (univariaten) Zufallsvariablen, d.h.  $\mathbf{x} = (x_1, \dots, x_p)^\top$ . Der Erwartungswert von  $x_i$ ,  $i = 1, \dots, p$  sei  $E(x_i) = \mu_i$ . Dann erhalten wir

$$E(\mathbf{x}) = \boldsymbol{\mu}, \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$$

als Vektor der Erwartungswerte bzw. Erwartungswertvektor.

- Gesucht: zusammenfassende Darstellung der Streuungsparameter (Varianzen und Kovarianzen) der  $p$  Zufallsvariablen. Es gilt

$$\begin{aligned} \text{Varianz:} \quad \text{Var}(x_i) &= E[\{x_i - E(x_i)\}^2] = E[\{x_i - E(x_i)\}\{x_i - E(x_i)\}] \\ \text{Kovarianz:} \quad \text{Cov}(x_i, x_j) &= E[\{x_i - E(x_i)\}\{x_j - E(x_j)\}] \end{aligned}$$

Allgemein gibt es  $p$  Varianzen und  $p(p-1)$  Kovarianzen, also  $p + p(p-1) = p^2$  Streuungsinformationen. Diese werden in der  $(p \times p)$ -Kovarianzmatrix (auch Varianz-Kovarianzmatrix) zusammengefasst:

$$\boldsymbol{\Sigma} := \text{Cov}(\mathbf{x}) = E[\{\mathbf{x} - E(\mathbf{x})\}\{\mathbf{x} - E(\mathbf{x})\}^\top]$$

Beispiel  $p = 2$ :

$$\begin{aligned} \boldsymbol{\Sigma} &= E \left[ \begin{pmatrix} x_1 - E(x_1) \\ x_2 - E(x_2) \end{pmatrix} \begin{pmatrix} x_1 - E(x_1) & x_2 - E(x_2) \end{pmatrix} \right] \\ &= E \left[ \begin{pmatrix} \{x_1 - E(x_1)\}\{x_1 - E(x_1)\} & \{x_1 - E(x_1)\}\{x_2 - E(x_2)\} \\ \{x_2 - E(x_2)\}\{x_1 - E(x_1)\} & \{x_2 - E(x_2)\}\{x_2 - E(x_2)\} \end{pmatrix} \right] \\ &= \begin{pmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) \end{pmatrix}. \end{aligned}$$

- Eigenschaften von  $\boldsymbol{\Sigma}$ :
  - quadratisch
  - symmetrisch
  - positiv-semidefinit (Zur Erinnerung: Matrix  $A$  ist positiv-semidefinit  $\Leftrightarrow \mathbf{x}^\top A \mathbf{x} \geq 0, \forall \mathbf{x} \neq \mathbf{0}$ )

#### Multivariate Normalverteilung

In den Abbildungen 1 bis 4 sind Dichten zweidimensionaler Normalverteilungen dargestellt.

Der allgemeine Fall:

$$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Bemerkung: Für unabhängige Zufallsvariablen ist  $\boldsymbol{\Sigma}$  eine Diagonalmatrix.

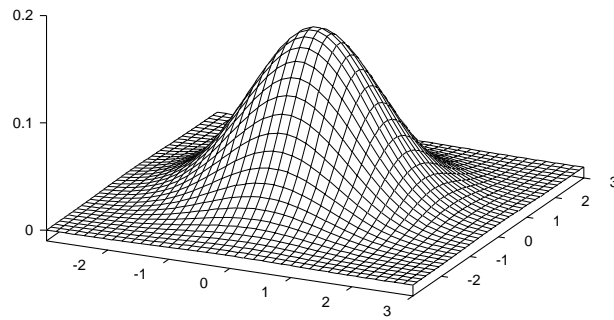


Abbildung 1: Zweidimensionale Normalverteilungsdichte für unkorrelierte Merkmale,  $\rho = 0$ , mit  $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1.0$

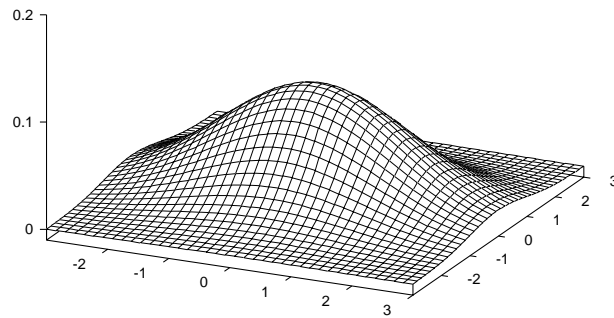


Abbildung 2: Zweidimensionale Normalverteilungsdichte für unkorrelierte Merkmale,  $\rho = 0$ , mit  $\mu_1 = \mu_2 = 0, \sigma_1 = 1.5, \sigma_2 = 1.0$

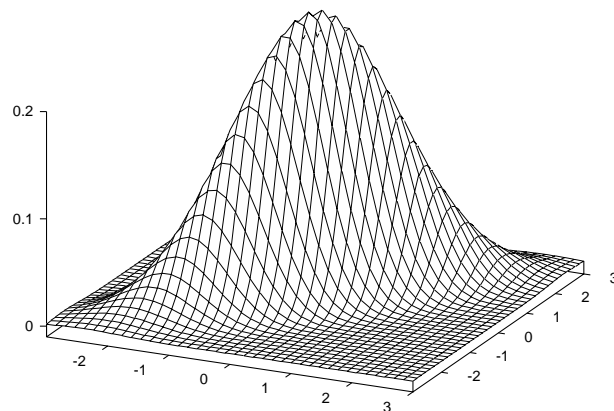


Abbildung 3: Zweidimensionale Normalverteilungsdichte,  $\rho = 0.8, \mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1.0$

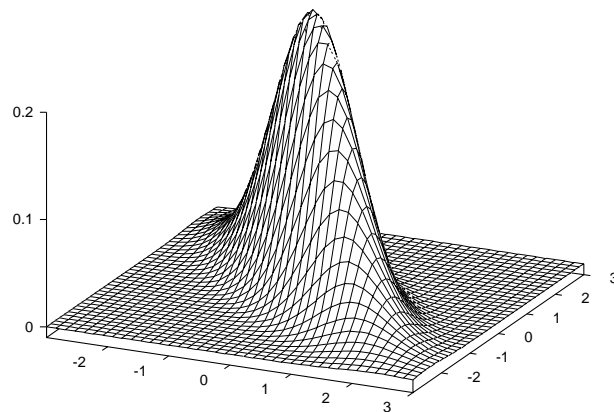


Abbildung 4: Zweidimensionale Normalverteilungsdichte,  $\rho = -0.8, \mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1.0$

## 4.2 Das multiple lineare Regressionsmodell

### 4.2.1 Modellformulierung

- Modell:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

mit

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

wobei

- $\mathbf{y}$ : abhängige Variable
- $\mathbf{X}$ : Regressormatrix
- $\boldsymbol{\epsilon}$ : Fehlerterm
- $\boldsymbol{\beta}$ : unbekannter Parametervektor
- Annahmen:
  - $\boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma^2 \mathbf{I}_n)$ , d.h. kein systematischer Fehler, die Fehlerterme sind unkorreliert und besitzen alle dieselbe Varianz (Homoskedastizität)
  - $\mathbf{X}$  deterministisch

### 4.2.2 KQ-Schätzung

- Prinzip der KQ-Schätzung: Minimiere die Summe der quadrierten Fehler, d.h.

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \epsilon_i^2 = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Alternativ gibt es folgende zwei Ansätze:

- Minimiere Summe der Fehler  $\implies$  Problem: positive und negative Fehler können sich aufheben. Dadurch kann es mehr als eine Lösung des Minimierungsproblems geben.
- Minimiere Summe der absoluten Fehler  $\implies$  Bestimmung der Lösung ist analytisch umständlich (vgl. Median)

**Satz 4.1.** Die KQ-Schätzer der unbekannt Parameter  $\boldsymbol{\beta}$  und  $\sigma^2$  lauten

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad \text{und} \quad \hat{\sigma} = \frac{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}{n - p - 1} = \frac{\mathbf{y}^\top \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{y}}{n - p - 1},$$

mit  $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ , falls  $\mathbf{X}$  vollen Rang  $p + 1$  besitzt..

*Beweis.* Zunächst zeigen wir, dass  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ .

- Nach dem Ansatz der KQ-Schätzung ist folgende Zielfunktion zu minimieren

$$S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}.$$

Notwendige Bedingung für ein Minimum ist, dass der Gradient gleich dem Nullvektor ist

$$\begin{aligned} \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} \stackrel{!}{=} \mathbf{0} \\ &\Leftrightarrow \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y} \\ &\stackrel{rg(\mathbf{X})=p+1}{\Leftrightarrow} \hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \end{aligned}$$

Hinreichende Bedingung für ein Minimum: Die Hesse-Matrix (Matrix der partiellen Ableitungen 2. Ordnung) muss positiv-semidefinit sein. In unserem Fall erhalten wir

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X} \geq 0, \quad \text{da } \mathbf{X}^\top \mathbf{X} \text{ positiv-semidefinit ist,}$$

d.h.  $\hat{\boldsymbol{\beta}}$  ist tatsächlich ein Minimum.

- Als nächstes betrachten wir die Herleitung des Schätzers für  $\sigma^2$ . Es gilt  $\hat{\sigma}^2 = \text{Var}(\epsilon_i) = \frac{1}{n} \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}$ . Problem: Der Vektor  $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$  der wahren Residuen ist unbekannt (da der wahre Koeffizientenvektor  $\boldsymbol{\beta}$  unbekannt ist). Lösung: Wir verwenden den Vektor  $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$  der geschätzten Residuen. Wie wir später sehen werden, ist

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}{n - p - 1}$$

ein erwartungstreuer Schätzer von  $\sigma^2$ . Nun ist noch zu zeigen, dass

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}{n - p - 1} \stackrel{(1)}{=} \frac{\mathbf{y}^\top \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{y}}{n - p - 1}.$$

Begründung für die Gültigkeit von Relation (1):

$$\begin{aligned} \hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y})^\top (\mathbf{y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}) \\ &= \mathbf{y}^\top (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)^\top (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{y} \\ &= \mathbf{y}^\top (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{y} \\ &= \mathbf{y}^\top \mathbf{y} - \underbrace{\mathbf{y}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}}_{\hat{\boldsymbol{\beta}}^\top}. \end{aligned}$$

□

- Eigenschaften:

- $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$ .
- KQ-Schätzer ist BLUE (bester linearer unverzerrter Schätzer nach dem Gauss-Markow-Theorem)
- $E(\hat{\sigma}^2) = \sigma^2$

*Beweis.* (i) Für den KQ-Schätzer erhalten wir

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon} \\ (\text{bzw. } \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon}). \end{aligned}$$

Damit können wir den Erwartungswertvektor und die Kovarianzmatrix des KQ-Schätzers berechnen:

$$E(\hat{\beta}) = \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E(\epsilon) = \beta$$

und

$$\begin{aligned} \text{Cov}(\hat{\beta}) &= E((\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^\top) \\ &= E((\hat{\beta} - \beta)(\hat{\beta} - \beta)^\top) \\ &= E((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \epsilon \epsilon^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}) \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E(\epsilon \epsilon^\top) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \\ &= \sigma^2 \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}}_{=\mathbf{I}_{p+1}} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \end{aligned}$$

Damit erhalten wir

$$\hat{\beta} \sim (\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}).$$

(ii) siehe Vorlesung „Lineare Modelle“.

(iii) Wie wir in Satz 4.1 gesehen haben, verwenden wir  $\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}$  für  $\hat{\sigma}^2$ . Es gilt

$$\begin{aligned} \hat{\epsilon} &= \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\ &= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{y} \\ &= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) (\mathbf{X}\beta + \epsilon) \\ &= \mathbf{X}\beta - \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}}_{=\mathbf{I}_{p+1}} \beta + (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \epsilon \\ &= \underbrace{\mathbf{X}\beta - \mathbf{X}\beta}_{=0} + (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \epsilon \quad (\text{lineare Funktion von } \epsilon) \\ &= \mathbf{M}\epsilon, \end{aligned}$$

wobei  $\mathbf{M} := (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)$  eine (deterministische) symmetrische und idempotente (d.h.  $\mathbf{M}\mathbf{M}^\top = \mathbf{M}^\top \mathbf{M} = \mathbf{M}^2 = \mathbf{M}$ ) Matrix ist, d.h. wir können schreiben

$$\hat{\epsilon}^\top \hat{\epsilon} = \epsilon^\top \mathbf{M}^\top \mathbf{M} \epsilon = \epsilon^\top \mathbf{M} \epsilon,$$

d.h. wir erhalten eine quadratische Form in  $\epsilon$ , mit anderen Worten einen Skalar.

Mit Hilfe des Spuroperators  $\text{tr}$  erhalten wir

$$\begin{aligned} E(\hat{\epsilon}^\top \hat{\epsilon}) &= E(\epsilon^\top \mathbf{M} \epsilon) \\ &= E[\text{tr}(\epsilon^\top \mathbf{M} \epsilon)] \quad (\epsilon^\top \mathbf{M} \epsilon \text{ ist ein Skalar!}) \\ &= E[\text{tr}(\mathbf{M} \epsilon \epsilon^\top)] \\ &= \text{tr}[\mathbf{M} E(\epsilon \epsilon^\top)] \\ &= \text{tr}[\mathbf{M} \sigma^2 \mathbf{I}_n] \\ &= \sigma^2 \text{tr}(\mathbf{M}) \\ &= \sigma^2 \text{tr}(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \\ &= \sigma^2 [\text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)] \quad (\text{verwende } \text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})) \\ &= \sigma^2 (n - p - 1) \end{aligned}$$

und damit  $E \left[ \frac{\hat{\epsilon}^\top \hat{\epsilon}}{n-p-1} \right] = \sigma^2$ .

□

### 4.2.3 Bestimmtheitsmaß

**Definition 4.2.** Das Bestimmtheitsmaß

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

misst den Anteil der durch das Modell erklärten Variation in  $y$  an der Gesamtvariation von  $y$ .

- Eigenschaften

(i)  $R^2 \stackrel{(2)}{=} \frac{\hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{y} - n\bar{y}^2}{\mathbf{y}^\top \mathbf{y} - n\bar{y}^2}$

(ii)  $0 \leq R^2 \leq 1$

(iii)  $R^2$  steigt automatisch mit  $p$

Begründung für Relation (2):

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\mathbf{y}}^\top \hat{\mathbf{y}} - n\bar{y}^2 = \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} - n\bar{y}^2 = \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} - n\bar{y}^2 = \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{y} - n\bar{y}^2$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{y}^\top \mathbf{y} - n\bar{y}^2$$

**Definition 4.3.** Das korrigierte Bestimmtheitsmaß  $\bar{R}^2$  ist definiert als

$$\bar{R}^2 := 1 - \left( \frac{n-1}{n-p-1} \right) (1 - R^2).$$

### 4.2.4 Prognose

- Sei  $\mathbf{y}_0$  der  $T_0$ -dimensionale Vektor der zu prognostizierenden abhängigen Variablen und  $\mathbf{X}_0$  die zugehörige Matrix der erklärenden Variablen. Dann gilt

$$\mathbf{y}_0 = \mathbf{X}_0 \boldsymbol{\beta} + \boldsymbol{\epsilon}_0$$

bzw. als Schätzung:

$$\hat{\mathbf{y}}_0 = \mathbf{X}_0 \hat{\boldsymbol{\beta}}$$

- Eigenschaften:

(i)  $E(\hat{\mathbf{y}}_0 - \mathbf{y}_0) = 0$  (Prognosefehler ist null)

(ii)  $E((\hat{\mathbf{y}}_0 - \mathbf{y}_0)^\top (\hat{\mathbf{y}}_0 - \mathbf{y}_0)) = \sigma^2 (\mathbf{X}_0 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_0^\top + \mathbf{I}_{T_0})$  (Prognosefehlervarianz)

- Prognose: Zur Erinnerung, es gilt  $\hat{\boldsymbol{\beta}} \sim (\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$  und damit

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0^\top \hat{\boldsymbol{\beta}} \sim (\mathbf{x}_0^\top \boldsymbol{\beta}, \sigma^2 \mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0).$$

Der wahre Wert lautet  $y_0 = \mathbf{x}_0^\top \boldsymbol{\beta} + \epsilon_0$ . Für den Prognosefehler  $\hat{y}_0 - y_0$  erhält man

$$\begin{aligned} E(\hat{y}_0 - y_0) &= E(\mathbf{x}_0^\top \hat{\boldsymbol{\beta}} - \mathbf{x}_0^\top \boldsymbol{\beta} - \epsilon_0) \\ &= E(\mathbf{x}_0^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \epsilon_0) \\ &= \mathbf{x}_0^\top E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - E(\epsilon_0) = 0. \end{aligned}$$

Für die Prognosefehlervarianz erhält man

$$\begin{aligned} E((\hat{y}_0 - y_0)^2) &= E((\mathbf{x}_0^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \epsilon_0)^2) \\ &= \mathbf{x}_0^\top E((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top) \mathbf{x}_0 + E(\epsilon_0^2) - 2\mathbf{x}_0^\top E(\epsilon_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\ &= \sigma^2 (\mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0 + 1). \end{aligned}$$

### 4.2.5 Hypothesentests

- Zusätzliche Annahme (für Tests und Konfidenzintervalle):  $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Damit ergibt sich

$$\begin{aligned} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} &\sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \\ \hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} &\sim N_{p+1}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}) = N_{p+1}(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}). \end{aligned}$$

Entsprechend gilt für die  $i$ -te Komponente des  $\hat{\boldsymbol{\beta}}$ -Vektors

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 (\mathbf{X}^\top \mathbf{X})_{i+1, i+1}^{-1})$$

bzw.

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 (\mathbf{X}^\top \mathbf{X})_{i+1, i+1}^{-1}}} \sim N(0, 1) \quad (\implies \text{Gauß-Test})$$

- Problem: Wir kennen  $\sigma^2$  nicht. Wir müssen daher  $\hat{\sigma}^2$  verwenden. Da  $\hat{\sigma}^2$  eine Zufallsgröße ist (hängt von den geschätzten Residuen ab, die wiederum vom KQ-Schätzer  $\hat{\boldsymbol{\beta}}$  abhängen, welcher wiederum von der Zufallsgröße  $\mathbf{y}$  abhängt!), stellt sich die Frage nach dessen Verteilung.
- Ansatz: Wir hatten gesehen, dass

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-p-1} = \frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{n-p-1}$$

gilt. Die neue Modellannahme lautet  $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Diese können wir folgendermaßen umschreiben

$$\frac{\boldsymbol{\epsilon}}{\sigma} \sim N_n(\mathbf{0}, \mathbf{I}_n) \Leftrightarrow \frac{\epsilon_i}{\sigma} \sim N(0, 1) \quad \text{für alle } i = 1, \dots, n.$$

Da das Quadrat einer standardnormalverteilten Zufallsgröße  $\chi^2$ -verteilt ist und die Summe von  $n$  quadrierten unabhängigen standardnormalverteilten Zufallsvariablen  $\chi^2$ -verteilt mit  $n$  Freiheitsgraden ist, erhalten wir

$$\frac{\epsilon_i^2}{\sigma^2} \sim \chi_1^2 \quad \implies \quad \frac{1}{\sigma^2} \sum_{i=1}^n \epsilon_i^2 = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}}{\sigma^2} \sim \chi_n^2.$$

Es gilt (ohne Beweis): Ist  $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$  und  $\mathbf{M}$  symmetrisch und idempotent mit  $rg(\mathbf{M}) = n-p-1$ , dann ist

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sigma^2} \sim \chi_{n-p-1}^2.$$

Damit erhalten wir

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-p-1} \frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sigma^2}$$

und

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-p-1} \chi_{n-p-1}^2 \quad \text{bzw.} \quad \frac{\hat{\sigma}^2 (n-p-1)}{\sigma^2} \sim \chi_{n-p-1}^2.$$

- Transformation multivariat normalverteilter Zufallsgrößen
  - Wiederholung: Im univariaten Fall gilt

$$\text{Ist } X \sim N(\mu, \sigma^2), \text{ dann ist } aX \sim N(a\mu, a^2\sigma^2) \quad (a \text{ ist eine Konstante}).$$

- Im multivariaten Fall gilt

$$\text{Ist } \mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ dann ist } \mathbf{A}\mathbf{x} \sim N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top),$$

wobei  $\mathbf{A}$  eine deterministische  $(q \times p)$ -dimensionale Matrix ist.

- Betrachte folgende Relation mit  $\mathbf{R}_1 := (r_0, r_1, \dots, r_p)$  als Zeilenvektor

$$\mathbf{R}_1 \boldsymbol{\beta} = r.$$

$\mathbf{R}_1 \boldsymbol{\beta}$  ist eine Linearkombination der Parameter. Es gilt

$$\mathbf{R}_1 \hat{\boldsymbol{\beta}} \sim N(\underbrace{\mathbf{R}_1 \boldsymbol{\beta}}_{(1 \times 1)}, \underbrace{\sigma^2 \mathbf{R}_1 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1^\top}_{(1 \times 1)}) \iff \frac{\mathbf{R}_1 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma \sqrt{\mathbf{R}_1 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1}} \sim N(0, 1).$$

- Problem:  $\sigma^2$  ist unbekannt. Lösung: Ersetzen durch  $\hat{\sigma}^2$ .
- Wiederholung: Definition der  $t$ -Verteilung: Seien  $Z \sim N(0, 1)$  und  $X \sim \chi_k^2$  unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{Z}{\sqrt{\frac{X}{k}}}$$

$t$ -verteilt mit  $k$  Freiheitsgraden.

- Es gilt

$$\frac{\mathbf{R}_1 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma} \sqrt{\mathbf{R}_1 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1}} = \frac{\frac{\mathbf{R}_1 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma \sqrt{\mathbf{R}_1 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1}}}{\sqrt{\frac{\hat{\sigma}^2 (n-p-1)}{\sigma^2 (n-p-1)}}} \sim t_{n-p-1}.$$

- Spezialfall: Ist  $\mathbf{R}_1 = (0, \dots, 0, 1, 0, \dots, 0)$  mit einer eins an der  $i$ -ten Stelle, dann gilt  $\mathbf{R}_1 \boldsymbol{\beta} = \beta_i$ . Wir erhalten

$$\frac{\mathbf{R}_1 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma} \sqrt{\mathbf{R}_1 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1}} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma} \sqrt{a_{i+1, i+1}}}, \quad \text{mit } a_{i+1, i+1} = (\mathbf{X}^\top \mathbf{X})_{i+1, i+1}^{-1} \quad ((i+1)\text{-tes Diagonalelement}).$$

- $(1 - \alpha)$ -Konfidenzintervalle für  $\mathbf{R}_1 \boldsymbol{\beta}$

$$\mathbf{R}_1 \hat{\boldsymbol{\beta}} \pm t_{1-\alpha/2, n-p-1} \hat{\sigma} \sqrt{\mathbf{R}_1 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1}.$$

- $(1 - \alpha)$ -Konfidenzintervall für Prognose  $y_0$ . Dabei sei  $y_0$  die abhängige Variable im Prognosezeitpunkt (unbekannt) und  $\mathbf{x}_0$  der Vektor der Ausprägungen der erklärenden Variablen im Prognosezeitpunkt (bekannt). Unter NV-Annahme gilt

$$\hat{y}_0 - y_0 \sim N(0, \sigma^2 (\mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0 + 1)) \iff \frac{\hat{y}_0 - y_0}{\sigma \sqrt{\mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0 + 1}} \sim N(0, 1).$$

Schätzung der Prognosefehlervarianz mit Hilfe von  $\hat{\sigma}^2$ . Damit erhalten wir dann

$$\frac{\hat{y}_0 - y_0}{\hat{\sigma} \sqrt{\mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0 + 1}} = \frac{\frac{\hat{y}_0 - y_0}{\sigma \sqrt{\mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0 + 1}}}{\sqrt{\frac{\hat{\sigma}^2 (n-p-1)}{\sigma^2 (n-p-1)}}} \sim t_{n-p-1}$$

Konfidenzintervall für  $y_0$ :

$$\mathbf{x}_0^\top \hat{\boldsymbol{\beta}} \pm t_{1-\alpha/2, n-p-1} \hat{\sigma} \sqrt{\mathbf{x}_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_0 + 1}$$

- Hypothesentests:

$$(i) H_0 : \mathbf{R}_1 \boldsymbol{\beta} = r \quad \text{vs.} \quad H_1 : \mathbf{R}_1 \boldsymbol{\beta} \neq r,$$



- (ii)  $H_0 : \mathbf{R}_1\boldsymbol{\beta} \geq r$  vs.  $H_1 : \mathbf{R}_1\boldsymbol{\beta} < r$ ,  
 (iii)  $H_0 : \mathbf{R}_1\boldsymbol{\beta} \leq r$  vs.  $H_1 : \mathbf{R}_1\boldsymbol{\beta} > r$ .

Unter  $H_0$  gilt

$$\mathbf{R}_1\hat{\boldsymbol{\beta}} \stackrel{H_0}{\sim} N(r, \sigma^2 \mathbf{R}_1(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1^\top).$$

Als Teststatistik (bei unbekanntem  $\sigma^2$ ) erhalten wir damit

$$T = \frac{\mathbf{R}_1\hat{\boldsymbol{\beta}} - r}{\hat{\sigma} \sqrt{\mathbf{R}_1(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}_1^\top}} \sim t_{n-p-1}.$$

Die Ablehnungsbereiche lauten

- (i)  $|T| > t_{1-\alpha/2, n-p-1}$ ,  
 (ii)  $T < -t_{1-\alpha, n-p-1}$ ,  
 (iii)  $T > t_{1-\alpha, n-p-1}$ .
- Overall-F-Test (Goodness-of-fit-Test): Die Nullhypothese lautet

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

d.h. kein Regressor besitzt einen signifikanten Erklärungsbeitrag zur Variation von  $y$ , gegen

$$H_1 : \beta_j \neq 0 \quad \text{für mindestens ein } j \in \{1, \dots, p\}.$$

Die Teststatistik lautet

$$F = \frac{R^2}{1 - R^2} \frac{n - p - 1}{p}.$$

Unter  $H_0$  gilt

$$F \stackrel{H_0}{\sim} F_{p, n-p-1}.$$

Der Ablehnungsbereich lautet

$$C = \{F : F > F_{1-\alpha, p, n-p-1}\}.$$

#### 4.2.6 Kodierung von Einflussgrößen

- Wir betrachten das Modell

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

- $x_i$  sei ein metrischer Regressor (z.B. Alter),  $y$  sei die Reaktionszeit

- Interpretation von  $\beta_i = 20 \frac{\text{msec}}{\text{Jahr}}$  :  $E(y|x_i)$  ändert sich um 20, wenn sich  $x_i$  um eine Einheit ändert
- Begründung:

$$E(y|x_1, \dots, x_i + 1, \dots, x_p) - E(y|x_1, \dots, x_i, \dots, x_p) = (x_i + 1)\beta_i - x_i\beta_i = \beta_i$$

Die anderen Regressoren seien dabei konstant (fest)!

- Beispiele:

- (i)  $E(\text{Nettomiete}) = \beta_0 + \# \text{Räume} \cdot \beta_A$ ,  $\beta_A = 88.58 \text{ Euro/Raum}$ ,  
 (ii)  $E(\text{Nettomiete}) = \beta_0 + \# \text{Räume} \cdot \beta_A + \text{Wohnfläche in } m^2 \cdot \beta_W$ ,  
 $\beta_A = -41.62 \text{ [Euro/Raum]}$ ,  $\beta_W = 6.135 \text{ [Euro/qm]}$ ,

- Nun sei  $x_1$  binär, z.B.

$$x_1 = \begin{cases} 1, & \text{männlich} \\ 0, & \text{weiblich} \end{cases}$$

Bei der naiven Anwendung des linearen Regressionsmodells würden wir folgendes erhalten

$$E(\text{Reaktionszeit}) = \beta_0 + x_1\beta_1, \quad \beta_1 = 30.2 \text{ [sek]}$$

- Frage: Ist  $\beta_1$  interpretierbar?
- Antwort: Ja. Männer sind um 30.2 Sekunden langsamer als Frauen.  $\beta_1$  entspricht der Veränderung, wenn  $x_i$  zu  $x_i + 1$  übergeht. Das entspricht hier dem Übergang von Kategorien. Daher ist eine Änderung von  $x_i$  nur nominal interpretierbar.
- Nun betrachten wir eine mehrkategoriale erklärende Variable  $M \in \{1, \dots, m\}$ . Zum Beispiel

$$M = \begin{cases} 1, & \text{morgens} \\ 2, & \text{mittags} \\ 3, & \text{abends} \end{cases}, \quad \text{oder} \quad M = \begin{cases} 1, & \text{Hamburg} \\ 2, & \text{München} \\ 3, & \text{Berlin} \end{cases}$$

- Frage: Ist ein Modell  $E(y|M) = \beta_0 + M\beta_M$  sinnvoll?
- Antwort: Nein, da  $M$  ein nominales Merkmal ist!
- Alternative: Kodierung der Kategorien (Einführung von dichotomen Kontrastvariablen)
- Zu den am häufigsten verwendeten Kodierungsansätzen zählen die *Dummy-Kodierung* und die *Effekt-Kodierung*. Bei der Dummy-Kodierung (auch 0 – 1-Kodierung) definiert man  $m$  Kontrastvariablen

$$x_i^M := \begin{cases} 1, & \text{wenn } M = i, \\ 0, & \text{sonst.} \end{cases}$$

Das naive lineare Regressionsmodell würde dann lauten:

$$E(y|M) = \beta_0 + x_1^M\beta_1 + \dots + x_m^M\beta_m.$$

- Problem: Wir erhalten  $m$  mögliche Erwartungswerte ( $E(y|M = 1), \dots, E(y|M = m)$ ), wir müssen aber  $(m+1)$  Parameter schätzen ( $\beta_0, \beta_1, \dots, \beta_m$ ). Das Modell wäre nicht identifizierbar!
- Lösung: Wir lassen einen Term weg, die sogenannte Referenzkategorie. Dabei gibt es unterschiedliche Konventionen, welcher Term als Referenzkategorie verwendet wird. Die geläufigsten sind

$$\beta_1 = 0 \iff M = 1 \text{ ist Referenzkategorie} \quad \text{oder} \quad \beta_m = 0 \iff M = m \text{ ist Referenzkategorie}$$

- Frage: Wie sieht es aus mit der Interpretation der Koeffizienten? Beispiel:  $m = 3$  wird als Referenzkategorie verwendet

$$E(y|M = 1) = \beta_0 + \beta_1$$

$$E(y|M = 2) = \beta_0 + \beta_2$$

$$E(y|M = 3) = \beta_0.$$

Damit erhalten wir  $\beta_i = E(y|M = i) - E(y|M = 3), i = 1, 2$ . Der Koeffizient  $\beta_i$  entspricht damit der Differenz zur Referenzkategorie  $m = 3$ .

- Die Alternative ist die Effekt-Kodierung. Wir betrachten das Modell

$$E(y|M = i) = \beta_0 + \beta_i, \quad i = 1, \dots, m.$$

- Wie wir bereits gesehen haben, haben wir einen Parameter zuviel im Modell, um Identifizierbarkeit zu gewährleisten. Die Idee der Effekt-Kodierung ist, diese Überparametrisierung durch eine zusätzliche Restriktion zu kompensieren, nämlich dass

$$\sum_{i=1}^m \beta_i = 0.$$

- Wenn wir die  $m$  Erwartungswerte  $E(y|M = 1) = \beta_0 + \beta_1$  bis  $E(y|M = m) = \beta_0 + \beta_m$  aufsummieren, dann erhalten wir unter Verwendung der Restriktion

$$\sum_{i=1}^m E(y|M = i) = m\beta_0 + \underbrace{\sum_{i=1}^m \beta_i}_{=0} \iff \beta_0 = \frac{1}{m} \sum_{i=1}^m E(y|M = i).$$

- Interpretation: Die Konstante  $\beta_0$  entspricht dem mittleren Effekt, d.h. der Mittelung über die Kategorien.
- Für  $\beta_i, i = 1, \dots, m$  gilt

$$\beta_i = E(y|M = i) - \beta_0,$$

d.h.  $\beta_i$  entspricht der Abweichung der  $i$ -ten Kategorie vom Mittel über alle Kategorien.

- äquivalente Darstellung der Effekt-Kodierung

$$x_i^M = \begin{cases} 1, & \text{wenn } M = i, \\ -1, & \text{wenn } M = m, \\ 0, & \text{sonst.} \end{cases}$$

Man erhält

$$E(y|M) = \beta_0 + x_1^M \beta_1 + \dots + x_{m-1}^M \beta_{m-1} \quad \text{und} \quad E(y|M = m) = \beta_0 - \sum_{i=1}^{m-1} \beta_i.$$

- Beispiel: Sei  $y$  der Mietpreis (in Euro pro Quadratmeter) und  $M$  die Wohngegend mit

$$M = \begin{cases} 1, & \text{Schwabing} \\ 2, & \text{Haidhausen} \\ 3, & \text{Neuperlach} \end{cases}$$

Folgende Daten sind gegeben:

$y$	10.81	7.95	8.50	8.25
$M$	1	3	2	3

Bei Dummy-Kodierung mit  $m = 3$  als Referenzkategorie erhalten wir folgendes Modell

$$E(y|M) = \beta_0 + x_1^M \beta_1 + x_2^M \beta_2 \quad \text{bzw.} \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

mit

$$\mathbf{y} = \begin{bmatrix} 10.81 \\ 7.95 \\ 8.50 \\ 8.25 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}.$$

Der KQ-Schätzer lautet

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 8.10 \\ 2.71 \\ 0.40 \end{bmatrix}$$

Dabei entspricht  $\hat{\beta}_0 = 8.10$  dem (durchschnittlichen) Mietpreis in Neuperlach (da Referenzkategorie!),  $\hat{\beta}_1 = 2.71$  der Differenz des durchschnittlichen Mietpreises zwischen Schwabing und Neuperlach und  $\hat{\beta}_2 = 0.40$  der Differenz zwischen Haidhausen und Neuperlach.

Bei Effekt-Kodierung mit  $m = 3$  als Referenzkategorie erhalten wir das Modell  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  mit

$$\mathbf{y} = \begin{bmatrix} 10.81 \\ 7.95 \\ 8.50 \\ 8.25 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}.$$

und als KQ-Schätzer

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 9.1367 \\ 1.6733 \\ -0.6367 \end{bmatrix}$$

sowie für  $\hat{E}(y|M=3) = \hat{\beta}_0 - \sum_{i=1}^2 \hat{\beta}_i = 8.1001 = \hat{\beta}_0 + \hat{\beta}_3 \iff \hat{\beta}_3 = -1.0366$ .

Interpretation:  $\hat{\beta}_0 = 9.1367$  entspricht dem mittleren Mietpreis (über alle Kategorien).  $\hat{\beta}_1 = 1.6733$  entspricht der Differenz Schwabing minus mittlerer Mietpreis, d.h. der mittlere Mietpreis in Schwabing liegt um 1.6733 Euro pro qm über dem mittleren Mietpreis. Äquivalent lassen sich  $\hat{\beta}_2$  und  $\hat{\beta}_3$  interpretieren.

- Mischung aus metrischen und kategorialen erklärenden Variablen. Wir betrachten folgendes Beispiel: Sei  $y$  der Umsatz (in 1000 Euro),  $x_1$  das investierte Kapital (in 1000 Euro) und  $x_2$  die Werbestrategie mit

$$x_2 = \begin{cases} 1, & \text{Werbestrategie 1,} \\ 0, & \text{Werbestrategie 2.} \end{cases}$$

Das Modell lautet

$$E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Der KQ-Schätzer sei

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1000 \\ 10 \\ 500 \end{bmatrix}$$

Für  $x_2 = 1$  erhalten  $E(y|x_1, x_2 = 1) = \beta_0 + \beta_1 x_1 + \beta_2 = 1000 + 10x_1 + 500$ , für  $x_2 = 0$  erhalten wir  $E(y|x_1, x_2 = 0) = \beta_0 + \beta_1 x_1 = 1000 + 10x_1$ . Interpretation: Werbestrategie 1 führt zu einem um 500 000 Euro höheren Umsatzniveau als Werbestrategie 2.

#### 4.2.7 Ausblick

- Erweiterung des KQ-Schätzers auf verallgemeinerte Kovarianzmatrix  $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \sigma^2\Psi$
- KQ-Schätzung bei nicht-linearen Funktionen  $y = f(\mathbf{x}, \boldsymbol{\beta}) + \epsilon$ . Der Ansatz lautet dann

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))^2 \rightarrow \min_{\boldsymbol{\beta}}$$

- Nichtparametrische Verfahren:

$$E(y|X=x) = \int y f(y|x) dx = \int y \frac{f(x, y)}{f(x)} dy$$

Schätzung von  $f(x, y)$  und  $f(x)$  z.B. mit (zweidimensionalen) Kerndichteschätzern

- Erweiterung auf generalisierte Regressionsmodelle  $\implies$  Erweiterung der Normalverteilungsannahme auf Exponentialfamilien.