Kapitel 1 - Das einfache lineare Regressionsmodell

Einfaches lineares Regressionsmodell

Das einfache lineare Regressionsmodell hat die Form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

für ein festes numerisches x_i und $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Beachte, dass per Definition gilt $Y_i | x_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$

Kleinste Quadrate (KQ) Schätzer

Wir schätzen die Parameter (β_0, β_1) durch

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_i))^2 \qquad (1)$$

und nennen $(\hat{\beta}_0, \hat{\beta}_1)$ den KQ-Schätzer von (β_0, β_1) und $\hat{\varepsilon}_i := Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ die Residuen.

Existenz und Berechnung vom KQ Schätzer

Der KQ-Schätzer existiert und ist eindeutig, falls $\sum_{i=1}^{n} (x_i - \overline{x})^2 \neq 0$. Dieser lässt sich berechnen als

$$\begin{aligned} \hat{\beta}_1 &= \frac{S_{xY}}{S_x^2} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2} \\ \hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1 \overline{x}. \end{aligned}$$

Durch differenzieren von der Gleichung (1) erhält man $(\hat{\beta}_0, \hat{\beta}_1)$ als Lösung der Normalengleichungen

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i x_i = 0$$

Interpretation der Modellparameter

Für $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i = 1, \ldots, n$ mit $\mathbb{E}(Y_i | x_i) = \beta_0 + \beta_1 x_i$ gilt,

- wenn x um eine **Einheit** steigt, dann steigt Y **im Erwartungswert** um β_1 Einheiten.
- Es gilt $\beta_0 = \mathbb{E}(Y | X = 0)$.
- Der Parameter σ die erwartete Abweichung der Y_i-Werte von der Regressionsgerade an.

Eigenschaften des KQ-Schätzers

Gegeben dem einfachen linearen Modell, gilt für den KQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1)$

- Erwartungstreue: $\mathbb{E}(\hat{\beta}_0, \hat{\beta}_1) = (\beta_0, \beta_1)$.
- $\mathbb{V}(\hat{\beta}_1) = \frac{\sigma^2}{n S_x^2}$ und $\mathbb{V}(\hat{\beta}_0) = \sigma^2 (\frac{1}{n} + \frac{\overline{x}^2}{n S_x^2})$.
- $(\hat{\beta}_0, \hat{\beta}_1)$ ist der maximum-likelihood Schätzer.

Schätzer für σ^2

Gegeben dem einfachen linearen Modell mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, gilt

$$\hat{\sigma}^2 := \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

ist ein erwartungstreuer Schätzer von σ^2 und

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2.$$

Der KQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1)$ und der Schätzer $\hat{\sigma}^2$ sind stoch.unabhängig.

Konfidenzintervalle für β_0 und β_1

Gegeben dem einfachen linearen Modell mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, gilt für $\hat{\beta}_1$ und $\hat{\beta}_0$

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{n-2} \text{ mit } \hat{\sigma}_{\hat{\beta}_1} := \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}$$

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t_{n-2} \text{ mit } \hat{\sigma}_{\hat{\beta}_0} := \sqrt{\hat{\sigma}^2 \frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \overline{x})^2}}$$

Damit können wir Konfidenzintervalle zum Niveau $1 - \alpha$ für β_1 und β_0 erzeugen:

$$[\ \hat{\beta}_1 - \hat{\sigma}_{\hat{\beta}_1} t_{1-\alpha/2}(n-2); \hat{\beta}_1 + \hat{\sigma}_{\hat{\beta}_1} t_{1-\alpha/2}(n-2)]$$

$$[\ \hat{\beta}_0 - \hat{\sigma}_{\hat{\beta}_0} t_{1-\alpha/2}(n-2); \hat{\beta}_0 + \hat{\sigma}_{\hat{\beta}_0} t_{1-\alpha/2}(n-2)]$$

Quadratsummenzerlegung

Gegeben sei ein einfaches linearen Modell mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ und $\hat{Y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i$. Dann gilt

$$\underbrace{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}_{\text{SSE}} - \underbrace{\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2}_{\text{SSM}}.$$

SST(otal):	Gesamtstreuung von Y
SSE(rror):	Streuung der Residuen
SSM(odel):	Streuung, die das Modell erklärt

Bestimmtheitsmaß

Unter Verwendung der obigen Notation definieren wir das Bestimmtheitsmaß als

$$R^2 = \frac{\text{SSM}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}.$$

Es gilt

$$R^2 = r_{xY}^2 = \frac{S_{xY}}{S_x S_Y},$$

wobei r_{xY} der Bravais-Pearson Korrel.koeffizient ist.

Interpretation von R^2

- R^2 beschreibt den Anteil der Varianz von Y, die durch x erklärt wird.
- R ist invariant gegenüber linearen linearen Transformationen von x und Y.
- R ist symmetrisch bzgl. x und Y.
- ! R^2 hängt auch von der Streuung von x in der Stichprobe ab.

Prognosewert

Gegeben sei ein einfaches linearen Modell mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ und $\hat{Y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, \ldots, n$. Sei nun eine weitere Beobachtung x_{n+1} mit zugehörigem $Y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1}$ gegeben. Der Prognosewert von Y_{n+1} ist definiert als $\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$

Prognosefehler

Gegeben sei ein einfaches linearen Modell, sowie eine weitere Beobachtung x_{n+1} mit zugehörigem Y_{n+1} sowie der Prognosewert $\hat{Y}_{n+1}$. Dann gilt

$$\mathbb{E}(\hat{Y}_{n+1} - Y_{n+1}) = 0$$

$$\mathbb{V}(\hat{Y}_{n+1} - Y_{n+1}) = \sigma^2 \Big[1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \Big]$$

Prognoseintervall

Gegeben sei ein einfaches linearen Modell, sowie eine weitere Beobachtung x_{n+1} mit zugehörigem Y_{n+1} sowie der Prognosewert $\hat{Y}_{n+1}$. Dann können wir für Y_{n+1} ein Konfidenzintervall zum Niveau $1 - \alpha$ konstruieren:

$$[\, \hat{Y}_{n+1} - \hat{\sigma}_{\hat{Y}_{n+1}} t_{1-\alpha/2}(n-2); \hat{Y}_{n+1} + \hat{\sigma}_{\hat{Y}_{n+1}} t_{1-\alpha/2}(n-2)]$$

mit

$$\hat{\sigma}_{\hat{Y}_{n+1}} = \hat{\sigma}^2 \Big[1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \Big].$$

R-Code

```
# simuliere aus einfachem lin. Modell
beta0 <- 3
beta1 <- 1
sigma <- 2
x <- seq(from = 0, to = 10, by = 0.5)
e <- rnorm(length(x), sd = sigma)
y <- beta0 + beta1 * x + e
dat <- data.frame(x, y)

# Lineares Modell erzeugen
reg = lm(y ~ x, data = dat)
summary(reg)

# Konfidenzintervalle
confint(reg, level = 0.95)
```

Interpretation von transformierten Modellen

- Log-Log-Modell:

  $$\log(Y_i) = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i$$

  Wenn x_i um den Faktor a steigt, dann steigt Y_i im Erwartungswert um den Faktor $a^{\beta_1} = e^{\beta_1 log(a)}$.
  Alternativ: Wenn x_i um 1% steigt, dann steigt Y_i im Erwartungswert um $(e^{\beta_1 log(1.01)} - 1)\%$.

- Linear-Log-Modell:

  $$Y_i = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i$$

  Wenn x_i um $p\%$ steigt, dann steigt Y_i im Erwartungswert um $\beta_1 \cdot \log(1 + p)\%$.
  Alternativ: Wenn x_i um 1% steigt, dann steigt Y_i im Erwartungswert um approximativ β_1 Einheiten.

- Log-Linear-Modell:

  $$\log(Y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$$

  Wenn x_i um eine Einheit steigt, dann steigt Y_i im Erwartungswert um e^{β_1} Einheiten.

Vorlesung

R^2 ist abhängig von X. Das heißt über mehrere Studien hinweg, die das gleiche messen, ist R^2 nur vergleichbar, wenn auch X vergleichbar ist. Je sicherer wir mit unserem Schätzer sein wollen, desto höher sollten wir die Varianz von X einstellen. Gegeben, dass der Zusammenhang tatsächlich linear ist, würde eine höhere Varianz von X zu einer geringeren Varianz von $\hat{\beta}_1$ führen.

Im multiplen Reg.modell ist es KEINE Annahme, dass x_i, x_j unabhängig voneinander sind. Es wäre nur praktisch für die Interpretation der Effekte. Das „magische“ am multiplen Reg.modell ist, dass ich für verschiedene Größen kontrollieren/korrigieren kann.

Erwartungstreue gilt auch bei abhängigkeit und normalverteilt ist nicht nötig. Varianzformel benötigt unabhängigkeit.

Kapitel 2 - Das multiple lineare Regressionsmodell

Multiples lineares Regressionsmodell

Das multiple lineare Regressionsmodell hat die Form

$$Y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}}_{\boldsymbol{x}_i^\top = (1, x_{i1}, \ldots, x_{ip})} + \varepsilon_i; i = 1, \ldots, n$$

oder in Matrix-Vektor Notation: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Wir nehmen dabei an, dass $\boldsymbol{X} \in \mathbb{R}^{n \times (p+1)}$ eine feste Design-Matrix ist und dass $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Wir definieren außerdem $p' := p + 1$.

Kleinste Quadrate (KQ) Schätzer

Wir schätzen den Parameter(vektor) $\boldsymbol{\beta}$ durch

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \qquad (2)$$

und nennen $\hat{\boldsymbol{\beta}}$ den KQ-Schätzer von $\boldsymbol{\beta}$ und $\hat{\varepsilon}_i := Y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$ die Residuen.

Existenz und Berechnung vom KQ Schätzer

Der KQ-Schätzer existiert und ist eindeutig, falls $\boldsymbol{X}^\top \boldsymbol{X}$ invertierbar ist. Dieser lässt sich berechnen als

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$

Durch differenzieren von der Gleichung (2) erhält man $\hat{\boldsymbol{\beta}}$ als Lösung der Normalengleichung

$$\boldsymbol{X}^\top \hat{\varepsilon} = 0$$

Interpretation der Modellparameter

- Y_i hängt linear von $x_{i1}, \ldots, x_{in}$ ab.
- Steigt x_k um eine Einheit, so steigt Y (ceteris paribus) im Erwartungswert um β_k Einheiten, **wenn** alle anderen x-Variablen festgehalten werden.
- ! β_k charakterisiert den Einfluss von x_k unter Berücksichtigung der übrigen Variablen (Confounder-Korrektur). Das heißt, dass in einem einfachen linearen Regressionsmodell mit $Y_i = \beta_0 + \beta_k' x_{ik} + \varepsilon_i$ wäre im Allgemeinen $\beta_k' \neq \beta_k$.

Eigenschaften des KQ-Schätzers

Gegeben dem multiplen linearen Modell, gilt für den KQ-Schätzer $\hat{\boldsymbol{\beta}}$

- Erwartungstreue: $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
  ! Gilt auch ohne die Annahme $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, solange $\mathbb{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$
- $\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$.
  ! Gilt auch ohne die Annahme $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, solange $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$
- $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1})$

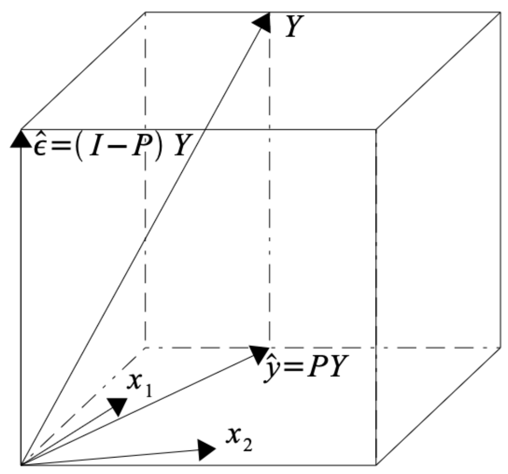
Hat-Matrix und Residualmatrix

Gegeben dem multiplen linearen Modell mit $\mathrm{rang}(\boldsymbol{X}) = p'$ gilt

$$\begin{aligned} \hat{\boldsymbol{Y}} &:= \boldsymbol{X} \underbrace{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}}_{\hat{\boldsymbol{\beta}}} \\ \boldsymbol{P} &:= \underbrace{\boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{n \times n} \\ \hat{\varepsilon} &= \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{P}) \boldsymbol{Y} \\ \boldsymbol{Q} &:= \boldsymbol{I} - \boldsymbol{P} \end{aligned}$$

$\boldsymbol{P}$ heißt Hat-Matrix und $\boldsymbol{Q}$ heißt Residualmatrix.

Geometrische Interpretation



Die KQ-Schätzung ist eine orthogonale Projektion von $\boldsymbol{Y}$ auf den von den $\boldsymbol{x}$-Vektoren aufgespannten Unterraum.

Eigenschaften von $\boldsymbol{P}$ und $\boldsymbol{Q}$

Die Hat-Matrix $\boldsymbol{P}$ und die Residualmatrix $\boldsymbol{Q}$ sind Projektionsmatrizen und zueinander orthogonal:

$$\begin{aligned} &\boldsymbol{P}^\top = \boldsymbol{P} \text{ und } \boldsymbol{P}^2 = \boldsymbol{P} \\ &\boldsymbol{Q}^\top = \boldsymbol{Q} \text{ und } \boldsymbol{Q}^2 = \boldsymbol{Q} \\ &\boldsymbol{P}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{P} = \boldsymbol{0}. \end{aligned}$$

Daraus folgt

$$\begin{aligned} &\mathbb{V}(\hat{\boldsymbol{Y}}) = \sigma^2 \boldsymbol{P} \\ &\mathbb{V}(\hat{\varepsilon}) = \sigma^2 \boldsymbol{Q}, \text{ da } \hat{\varepsilon} = \boldsymbol{Q}\boldsymbol{\varepsilon} \end{aligned}$$

Schätzer für σ^2

Gegeben dem multiplen linearen Modell, gilt

$$\hat{\sigma}^2 := \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{n - (p+1)} = \frac{1}{n - (p+1)} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

ist ein erwartungstreuer Schätzer von σ^2.
! Gilt auch ohne die Annahme $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, solange $\mathbb{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ und $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$

Kapitel 3 - Quadratsummenzerlegung und statistische Inferenz im multiplen linearen Regressionsmodell

Quadratsummenzerlegung

Gegeben sei das multiple lineare Regressionsmodell mit $\text{rang}(\boldsymbol{X}) = p'$. Dann gilt

$$\underbrace{(\boldsymbol{Y}-\overline{\boldsymbol{Y}})^\top(\boldsymbol{Y}-\overline{\boldsymbol{Y}})}_{SST} = \underbrace{(\boldsymbol{Y}-\hat{\boldsymbol{Y}})^\top(\boldsymbol{Y}-\hat{\boldsymbol{Y}})}_{SSE} + \underbrace{(\hat{\boldsymbol{Y}}-\overline{\boldsymbol{Y}})^\top(\hat{\boldsymbol{Y}}-\overline{\boldsymbol{Y}})}_{SSM}.$$

SST(otal):	Gesamt-Quadratsumme (korrigiert)
SSE(rror):	Fehler-Quadratsumme
SSM(odel):	Modell-Quadratsumme

Quadratsummenzerlegung ohne β_0

Gegeben sei das multiple lineare Regressionsmodell mit, aber ohne Absolutglied β_0. Dann gilt

$$\underbrace{\boldsymbol{Y}^\top\boldsymbol{Y}}_{SST^*} = \underbrace{(\boldsymbol{Y}-\hat{\boldsymbol{Y}})^\top(\boldsymbol{Y}-\hat{\boldsymbol{Y}})}_{SSE} + \underbrace{\hat{\boldsymbol{Y}}^\top\hat{\boldsymbol{Y}}}_{SSM^*}.$$

SST^*:	Gesamt-Quadratsumme (nicht korrigiert)
SSE:	Fehler-Quadratsumme (wie zuvor)
SSM^*:	Modell-Quadratsumme (nicht korrigiert)

Erwartungswerte der Quadratsummen

Gegeben sei das multiple lineare Regressionsmodell mit den üblichen Annahmen. Wir definieren

$$\boldsymbol{e} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \text{ und } \boldsymbol{P_e} = \boldsymbol{e}(\boldsymbol{e}^\top\boldsymbol{e})^{-1}\boldsymbol{e}^\top \text{ und } \boldsymbol{Q_e} = \boldsymbol{I} - \boldsymbol{P_e}.$$

Dann gilt

$$\begin{aligned}
\mathbb{E}(SST^*) &= \sigma^2 n + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} \\
\mathbb{E}(SST) &= \sigma^2(n-1) + \boldsymbol{\beta}^\top(\boldsymbol{Q_e}\boldsymbol{X})^\top(\boldsymbol{Q_e}\boldsymbol{X})\boldsymbol{\beta} \\
\mathbb{E}(SSE) &= \sigma^2(n-p') \\
\mathbb{E}(SSM^*) &= \sigma^2 p' + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} \\
\mathbb{E}(SSM^*) &= \sigma^2(p'-1) + \boldsymbol{\beta}^\top(\boldsymbol{Q_e}\boldsymbol{X})^\top(\boldsymbol{Q_e}\boldsymbol{X})\boldsymbol{\beta}
\end{aligned}$$

Wir können diese Eigenschaften zur Konstruktion von Tests verwenden. Es gilt nämlich unter anderem
$\boldsymbol{\beta} = 0 \implies \mathbb{E}(SST^*) = \sigma^2 n$
$\beta_1 = \cdots = \beta_p = 0 \implies \mathbb{E}(SSM) = \sigma^2(p'-1)$

Chi-Quadrat Verteilung

Sei $\boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{I})$, so heißt $\boldsymbol{W} = \boldsymbol{Z}^\top\boldsymbol{Z} = \sum_{i=1}^n Z_i^2$ (nicht-zentral) Chi-Quadrat-verteilt und wir schreiben

$$W \sim \chi^2(n, \delta).$$

Wir nennen n die Zahl der Freiheitsgrade und $\delta = \boldsymbol{\mu}^\top\boldsymbol{\mu}$ den Nicht-Zentralitätsparameter. Es gilt

$$\begin{aligned}
\mathbb{E}(W) &= n + \delta \\
\mathbb{V}(W) &= 2n + 4\delta
\end{aligned}$$

t-Verteilung

Seien $Z \sim \mathcal{N}(\delta, 1)$ und $W \sim \chi^2(n, 0)$ unabhängig. Dann heißt $T = \frac{Z}{\sqrt{\frac{W}{n}}}$ (nicht-zentral) t-verteilt mit n Freiheitsgraden und Nicht-Zentralitätsparameter δ und wir schreiben

$$T \sim t(n, \delta).$$

Es gilt

$$\mathbb{E}(T) = \delta\sqrt{\frac{n}{2}}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \text{ für } n > 1$$

F-Verteilung

Sei $W_1 \sim \chi^2(n_1, \delta)$ und $W_2 \sim \chi^2(n_2, 0)$ unabhängig. Dann heißt $X = \frac{W_1/n_1}{W_2/n_2}$ (nicht-zentral) F-verteilt mit n_1 und n_2 Freiheitsgraden und Nicht-Zentralitätsparameter δ und wir schreiben

$$X \sim F(n_1, n_2, \delta).$$

Es gilt

$$\mathbb{E}(X) = \frac{n_2 + \frac{n_2\delta}{n_1}}{n_2 - 2} \text{ für } n_2 > 2$$

Kapitel 4 - Diskrete Einflußgrößen

Kodierung

Sei C eine nominale Variable mit K Ausprägungen.

Dummy/Referenz-Kodierung:
Wir definieren K neue Variablen $Z_1, \ldots, Z_K$ als

$$Z_k(C) = \begin{cases} 1, & \text{falls } C = k \\ 0, & \text{sonst} \end{cases}$$

$Z_1, \ldots, Z_K$ sind abhängig, da $Z_K = 1 - \sum_{k=1}^{K-1} Z_k$

Effekt-Kodierung: Wir definieren $K-1$ neue Variablen $Z_1^e, \ldots, Z_{K-1}^e$ als

$$Z_k^e(C) = \begin{cases} 1, & \text{falls } C = k \\ -1, & \text{falls } C = K \\ 0, & \text{sonst} \end{cases}$$

Note: $Z_k(\boldsymbol{C}) = \begin{pmatrix} Z_k(C_1) \\ \vdots \\ Z_k(C_n) \end{pmatrix}$ und $Z_k^e(\boldsymbol{C}) = \begin{pmatrix} Z_k^e(C_1) \\ \vdots \\ Z_k^e(C_n) \end{pmatrix}$

Setup einfache Varianzanalyse

Im folgenden betrachten wir die einfache Varianzanalyse mit nur einer diskreten Einflußgröße $\boldsymbol{C} = \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix}$ mit K Ausprägungen. Sei n_k dabei die Anzahl der Beobachtungen mit $C_i = k$.

Mittelwertsmodell

Das Mittelwertsmodell ist gegeben durch

$$Y_{k_l} = \mu_k + \epsilon_{k_l} \quad l = 1, \ldots, n_k \quad k = 1, \ldots, K$$

oder in Matrix-Vektor Notation:

$$\boldsymbol{Y} = (Z_1(\boldsymbol{C}) \cdots Z_K(\boldsymbol{C})) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix} + \boldsymbol{\epsilon}$$

Bei dem Mittelwertsmodell gibt es keinen Intercept und die μ_k sind die Mittelwerte der k-ten Gruppe. Der Effekt der k-ten Gruppe ist also μ_k.

Mittelwertsmodell Beispiel

Für $K = 3$ Ausprägungen und $n_k = 2$ für alle $k = 1, 2, 3$ erhalten wir als Mittelwertsmodell:

$$\boldsymbol{Y} = \begin{pmatrix} Y_{1_1} \\ Y_{1_2} \\ Y_{2_1} \\ Y_{2_2} \\ Y_{3_1} \\ Y_{3_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1_1} \\ \varepsilon_{1_2} \\ \varepsilon_{2_1} \\ \varepsilon_{2_2} \\ \varepsilon_{3_1} \\ \varepsilon_{3_2} \end{pmatrix}$$

Modell mit Effekt-Kodierung

Das Modell mit Effekt-Kodierung ist gegeben durch

$$Y_{k_l} = \mu + \tau_k + \epsilon_{k_l}; \quad \tau_K = -\sum_{k=1}^{K-1} \tau_k$$

für $\quad l = 1, \ldots, n_k \quad k = 1, \ldots, K$ oder in Matrix-Vektor Notation:

$$\boldsymbol{Y} = (\boldsymbol{e} \;\; Z_1^e(\boldsymbol{C}) \cdots Z_{K-1}^e(\boldsymbol{C})) \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_{K-1} \end{pmatrix} + \boldsymbol{\epsilon}$$

Bei dem Modell mit Effekt-Kodierung gibt es einen Intercept μ und die τ_k sind die Abweichungen der k-ten Gruppe vom Gesamtmittelwert bzw. vom Intercept μ. Der Effekt der k-ten Gruppe ist also $\mu + \tau_k$.

Modell mit Effekt-Kodierung Beispiel

Für $K = 3$ Ausprägungen und $n_k = 2$ für alle $k = 1, 2, 3$ erhalten wir als Modell mit Effekt-Kodierung:

$$\boldsymbol{Y} = \begin{pmatrix} Y_{1_1} \\ Y_{1_2} \\ Y_{2_1} \\ Y_{2_2} \\ Y_{3_1} \\ Y_{3_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1_1} \\ \varepsilon_{1_2} \\ \varepsilon_{2_1} \\ \varepsilon_{2_2} \\ \varepsilon_{3_1} \\ \varepsilon_{3_2} \end{pmatrix}$$

Modell mit Referenz-Kodierung

Das Modell mit Referenz-Kodierung ist gegeben durch

$$Y_{k_l} = \mu_K + \tau_k + \epsilon_{k_l}; \quad \tau_K = 0$$

für $\quad l = 1, \ldots, n_k \quad k = 1, \ldots, K$ oder in Matrix-Vektor Notation:

$$\boldsymbol{Y} = (\boldsymbol{e} \;\; Z_1(\boldsymbol{C}) \cdots Z_{K-1}(\boldsymbol{C})) \begin{pmatrix} \mu_K \\ \tau_1 \\ \vdots \\ \tau_{K-1} \end{pmatrix} + \boldsymbol{\epsilon}$$

Beim Modell mit Referenz-Kodierung gibt es einen Intercept μ_K der den Mittelwert der K-ten Gruppe angibt und die τ_k sind die Abweichungen der k-ten Gruppe vom Mittelwert der K-ten Referenz-Gruppe. Der Effekt der k-ten Gruppe ist also $\mu_K + \tau_k$ für $k = 1, \ldots, K-1$ und μ_K für $k = K$.

Modell mit Referenz-Kodierung Beispiel

Für $K = 3$ Ausprägungen und $n_k = 2$ für alle $k = 1, 2, 3$ erhalten wir als Modell mit Referenz-Kodierung:

$$\boldsymbol{Y} = \begin{pmatrix} Y_{1_1} \\ Y_{1_2} \\ Y_{2_1} \\ Y_{2_2} \\ Y_{3_1} \\ Y_{3_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_3 \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1_1} \\ \varepsilon_{1_2} \\ \varepsilon_{2_1} \\ \varepsilon_{2_2} \\ \varepsilon_{3_1} \\ \varepsilon_{3_2} \end{pmatrix}$$

Bemerkungen-Kodierung

Alle Modellvarianten führen zur gleichen Modellanpassung (R^2). Die Parameter haben aber unterschiedliche Interpretationen. Parameter und deren Schätzer sind aber ineinander umrechenbar.
Wir können folgende Nullhypothese für den Effekt von C testen:

Mittelwertsmodell	$H_0 : \mu_1 = \cdots = \mu_K$
Effekt-Kodierung	$H_0 : \tau_1 = \cdots = \tau_{K-1} = 0$
Referenz-Kodierung	$H_0 : \tau_1 = \cdots = \tau_{K-1} = 0$

Setup zweifaktorielle Varianzanalyse

Im folgenden betrachten wir zwei diskrete Einflußgrößen $\boldsymbol{C} = \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix}$ und $\boldsymbol{D} = \begin{pmatrix} D_1 \\ \vdots \\ D_n \end{pmatrix}$ mit K_C bzw. K_D Ausprägungen. Sei $n_{k,l}$ dabei die Anzahl der Beobachtungen mit $C_i = k$ und $D_j = l$.

! Hier ist die Mittelwertsdarstellung bzw. das Mittelwertsmodell nicht möglich, da dieser davon abhängig ist, welche Variable zuerst kodiert wird.

Modell mit Effekt-Kodierung (mehrfaktoriell)

Das Modell mit Effekt-Kodierung ist gegeben durch

$$\boldsymbol{Y} = (\boldsymbol{e} \quad Z_1^e(\boldsymbol{C}) \cdots Z^e_{K_C-1}(\boldsymbol{C}) \quad Z_1^e(\boldsymbol{D}) \cdots Z^e_{K_D-1}(\boldsymbol{D})) \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_{K_C-1} \\ \gamma_1 \\ \vdots \\ \gamma_{K_D-1} \end{pmatrix} + \boldsymbol{\epsilon}$$

mit $\tau_{K_C} = -\sum_{k=1}^{K_C-1} \tau_k$ und $\gamma_{K_D} = -\sum_{k=1}^{K_D-1} \gamma_k$.

Bei dem Modell mit Effekt-Kodierung gibt es einen Intercept μ und die τ_k und γ_l sind die Abweichungen der Gruppe mit $C = k$ bzw. $D = l$ vom Gesamtmittelwert bzw. vom Intercept μ.

Modell mit Referenz-Kodierung (mehrfakt.)

Das Modell mit Referenz-Kodierung ist gegeben durch

$$\boldsymbol{Y} = (\boldsymbol{e} \quad Z_1(\boldsymbol{C}) \cdots Z_{K_C-1}(\boldsymbol{C}) \quad Z_1(\boldsymbol{D}) \cdots Z_{K_D-1}(\boldsymbol{D})) \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_{K_C-1} \\ \gamma_1 \\ \vdots \\ \gamma_{K_D-1} \end{pmatrix} + \boldsymbol{\epsilon}$$

mit $\tau_{K_C} = 0$ und $\gamma_{K_D} = 0$.

Bei dem Modell mit Referenz-Kodierung gibt es einen Intercept μ der den Mittelwert der Gruppe mit $C = K_C$ und $D = K_D$ angibt und die τ_k und γ_l sind die Abweichungen der Gruppe mit $C = k$ bzw. $D = l$ vom Mittelwert der Gruppe mit $C = K_C$ und $D = K_D$.

Kodierung Vergleich (mehrfaktoriell) Beispiel

Sei $K_C = 2$ und $K_D = 3$ mit $n_{k,l} = 2$ für alle $k = 1, 2, 3$ und $l = 1, 2$.
Dann erhalten wir als Designmatrix für das Modell mit

Effekt-Kodierung:

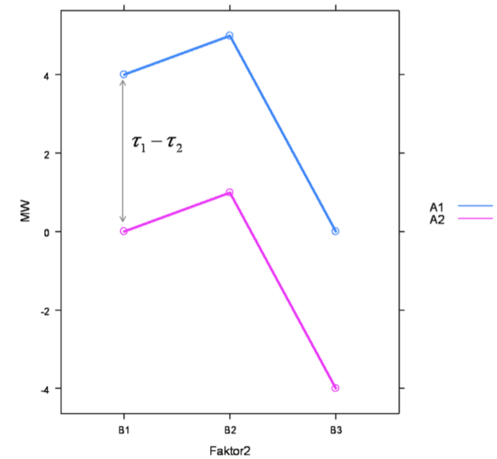
$$\boldsymbol{X} = (\boldsymbol{e}\ Z_1^e(\boldsymbol{C})\ Z_1^e(\boldsymbol{D})\ Z_2^e(\boldsymbol{D})) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix}$$

Referenz-Kodierung:

$$\boldsymbol{X} = (\boldsymbol{e}\ Z_1(\boldsymbol{C})\ Z_1(\boldsymbol{D})\ Z_2(\boldsymbol{D})) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Visualisierung Beispiel

Wir können die Effekte visualisieren, indem wir die Mittelwerte der Gruppen betrachten:



Note: „Faktor 2" ist hier D und „A1" und „A2" sind hier $C = 1$ und $C = 2$. Auf der y-Achse ist der Mittelwert der Gruppe dargestellt.

In beiden Fällen werden folgende Gleichungen erfüllt:

$$\begin{aligned} \widehat{\mu} + \widehat{\tau_1} + \widehat{\gamma_1} &= 4 & \widehat{\mu} + \widehat{\tau_2} + \widehat{\gamma_1} &= 0 \\ \widehat{\mu} + \widehat{\tau_1} + \widehat{\gamma_2} &= 5 & \widehat{\mu} + \widehat{\tau_2} + \widehat{\gamma_2} &= 1 \\ \widehat{\mu} + \widehat{\tau_1} + \widehat{\gamma_3} &= 0 & \widehat{\mu} + \widehat{\tau_2} + \widehat{\gamma_3} &= -4 \end{aligned}$$

Effekt-Kodierung:

$$\begin{aligned} \widehat{\mu} &= 1 & \widehat{\gamma_1} &= 1 \\ \widehat{\tau_1} &= 2 & \widehat{\gamma_2} &= 2 \\ \widehat{\tau_2} &= -2 & \widehat{\gamma_3} &= -3 \end{aligned}$$

! Der Verlauf für $C = 1$ und $C = K_C = 2$ ist parallel mit Abstand $\widehat{\tau_1} - \widehat{\tau_2}$.

Referenz-Kodierung:

$$\begin{aligned} \widehat{\mu} &= -4 & \widehat{\gamma_1} &= 4 \\ \widehat{\tau_1} &= 4 & \widehat{\gamma_2} &= 5 \\ \widehat{\tau_2} &= 0 & \widehat{\gamma_3} &= 0 \end{aligned}$$

! Der Verlauf für $C = 1$ und $C = K_C = 2$ ist parallel mit Abstand $\widehat{\tau_1}$. Man beachte auch, dass sich die Schätzer direkt in dem Plot ablesen lassen.

Interaktion

In den obigen Modellen haben wir die Interaktion zwischen den Einflußgrößen C und D nicht berücksichtigt. Interaktion bedeutet, dass der Effekt von C von D abhängt und umgekehrt.

Beispiele:

- Die Wirkung des Medikaments ist bei Männern anders als bei Frauen.
- Die Wirkung des Medikaments ist bei jungen Menschen anders als bei alten Menschen.
- Die Wirkung des Lesetrainings ist bei guten Schülern geringer als bei schwachen Schülern.

! Der Begriff Interaktion ist in anderen Fachbereichen auch bekannt als Moderation (Psychologie), Synergieefffekte (Wirtschaft).

Wir können die Interaktion zwischen C und D berücksichtigen, indem wir die Effekte von C und D nicht additiv, sondern multiplikativ betrachten.

Interaktionsmodell (Effekt-Kodierung)

In dem Modell mit Effekt-Kodierung und Interaktion ist die Designmatrix gegeben durch die Spalten der Designmatrix aus dem Modell ohne Interaktion $\boldsymbol{X}^e$ und zusätzlich die Spalten der Interaktionsterme:

$$\boldsymbol{Z}^e = (Z_1^e(\boldsymbol{C})Z_1^e(\boldsymbol{D}) \cdots Z_1^e(\boldsymbol{C})Z_{K_D-1}^e(\boldsymbol{D}) \cdots Z_{K_C-1}^e(\boldsymbol{C})Z_{K_D-1}^e(\boldsymbol{D}))$$

Die Designmatrix ist also gegeben durch $\begin{pmatrix}\boldsymbol{X}^e & \boldsymbol{Z}^e\end{pmatrix}$

Die Parameter sind gegeben durch $\mu \quad \tau_k \quad \gamma_l \quad (\tau\gamma)_{k,l}$ mit $k = 1, \ldots, K_C - 1 \quad l = 1, \ldots, K_D - 1$.

Die Modellgleichung ist gegeben durch

$$Y_{k,l} = \mu + \tau_k + \gamma_l + (\tau\gamma)_{k,l} + \epsilon_{k,l}$$

mit den Nebenbedingungen

$$\sum_{k=1}^{K_C-1} \tau_k = 0 \quad \text{und} \quad \sum_{l=1}^{K_D-1} \gamma_l = 0$$

sowie

$$\forall l : \sum_{k=1}^{K_C} (\tau\gamma)_{k,l} = 0 \quad \text{und} \quad \forall k : \sum_{l=1}^{K_D} (\tau\gamma)_{k,l} = 0$$

Interpretation Interaktion (Effekt-Kodierung)

- μ ist der Gesamtmittelwert.
- τ_k ist der Unterschied zwischen dem Mittelwert der Gruppe mit $C = k$ und dem Gesamtmittelwert.
- γ_l ist der Unterschied zwischen dem Mittelwert der Gruppe mit $D = l$ und dem Gesamtmittelwert.
- $(\tau\gamma)_{k,l}$ ist der Effekt der Interaktion auf die Basiseffekte τ_k und γ_l durch die Ausprägungen $C = k$ und $D = l$.

Interaktionsmodell (Referenz-Kodierung)

In dem Modell mit Referenz-Kodierung und Interaktion ist die Designmatrix gegeben durch die Spalten der Designmatrix aus dem Modell ohne Interaktion $\boldsymbol{X}$ und zusätzlich die Spalten der Interaktionsterme:

$$\boldsymbol{Z} = (Z_1(\boldsymbol{C})Z_1(\boldsymbol{D}) \cdots Z_1(\boldsymbol{C})Z_{K_D-1}(\boldsymbol{D}) \cdots Z_{K_C-1}(\boldsymbol{C})Z_{K_D-1}(\boldsymbol{D}))$$

Die Designmatrix ist also gegeben durch $\begin{pmatrix}\boldsymbol{X} & \boldsymbol{Z}\end{pmatrix}$.

Die Parameter sind gegeben durch $\mu \quad \tau_k \quad \gamma_l \quad (\tau\gamma)_{k,l}$ mit $k = 1, \ldots, K_C - 1 \quad l = 1, \ldots, K_D - 1$.

Die Modellgleichung ist gegeben durch

$$Y_{k,l} = \mu + \tau_k + \gamma_l + (\tau\gamma)_{k,l} + \epsilon_{k,l}$$

mit den Nebenbedingungen

$$\tau_{K_C} = 0 \quad \text{und} \quad \gamma_{K_D} = 0$$

sowie

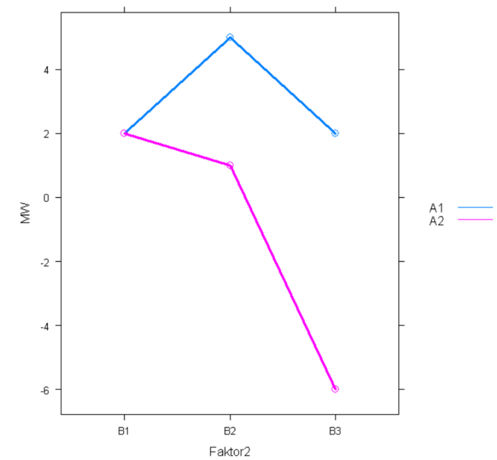
$$\forall l : (\tau\gamma)_{K_C,l} = 0 \quad \text{und} \quad \forall k : (\tau\gamma)_{k,K_D} = 0$$

Interpret. Interaktion (Referenz-Kodierung)

- τ_k ist der Unterschied zwischen dem Mittelwert der Gruppe mit $C = k$ zu der Referenzgruppe mit $C = K_C$.
- γ_l ist der Unterschied zwischen dem Mittelwert der Gruppe mit $D = l$ zu der Referenzgruppe mit $D = K_D$.
- $(\tau\gamma)_{k,l}$ ist der Effekt der Interaktion auf die Basiseffekte τ_k und γ_l zur jeweiligen Referenzgruppe durch die Ausprägungen $C = k$ und $D = l$.

Visualisierung Beispiel

Wir können die Effekte visualisieren, indem wir die Mittelwerte der Gruppen betrachten:



Note: „Faktor 2“ ist hier D und „A1“ und „A2“ sind hier $C = 1$ und $C = 2$. Auf der y-Achse ist der Mittelwert der Gruppe dargestellt.

In beiden Fällen werden folgende Gleichungen erfüllt:

$$\begin{aligned}
\widehat{\mu} + \widehat{\tau}_1 + \widehat{\gamma}_1 + \widehat{(\tau\gamma)}_{1,1} &= 2 \\
\widehat{\mu} + \widehat{\tau}_1 + \widehat{\gamma}_2 + \widehat{(\tau\gamma)}_{1,2} &= 5 \\
\widehat{\mu} + \widehat{\tau}_1 + \widehat{\gamma}_3 + \widehat{(\tau\gamma)}_{1,3} &= 2 \\
\widehat{\mu} + \widehat{\tau}_2 + \widehat{\gamma}_1 + \widehat{(\tau\gamma)}_{2,1} &= 2 \\
\widehat{\mu} + \widehat{\tau}_2 + \widehat{\gamma}_2 + \widehat{(\tau\gamma)}_{2,2} &= 1 \\
\widehat{\mu} + \widehat{\tau}_2 + \widehat{\gamma}_3 + \widehat{(\tau\gamma)}_{2,3} &= -6
\end{aligned}$$

Effekt-Kodierung:

$$\begin{array}{llll}
\widehat{\mu} = 1 & \widehat{\gamma}_1 = 1 & \widehat{(\tau\gamma)}_{1,1} = -2 & \widehat{(\tau\gamma)}_{2,1} = 2 \\
\widehat{\tau}_1 = 2 & \widehat{\gamma}_2 = 2 & \widehat{(\tau\gamma)}_{1,2} = 0 & \widehat{(\tau\gamma)}_{2,2} = 0 \\
\widehat{\tau}_2 = -2 & \widehat{\gamma}_3 = -3 & \widehat{(\tau\gamma)}_{1,3} = 2 & \widehat{(\tau\gamma)}_{2,3} = -2
\end{array}$$

Referenz-Kodierung:

$$\begin{array}{llll}
\widehat{\mu} = -6 & \widehat{\gamma}_1 = 8 & \widehat{(\tau\gamma)}_{1,1} = -2 & \widehat{(\tau\gamma)}_{2,1} = 0 \\
\widehat{\tau}_1 = 4 & \widehat{\gamma}_2 = 7 & \widehat{(\tau\gamma)}_{1,2} = 1 & \widehat{(\tau\gamma)}_{2,2} = 0 \\
\widehat{\tau}_2 = 0 & \widehat{\gamma}_3 = 0 & \widehat{(\tau\gamma)}_{1,3} = 0 & \widehat{(\tau\gamma)}_{2,3} = 0
\end{array}$$