Statistische Modellierung III

-Das allgemeine lineare Modell-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Das allgemeine lineare Modell

Allgemeines lin. Modell - Setup

- Betrachten wieder Beobachtungen $(Y_i, \mathbf{x}_i)$, $i = 1, \dots, n$, wobei $\mathbf{x}_i \in \mathbb{R}^k$ fixe Kovariablen und die Zielvariablen Y_i wieder Zufallsvariablen sind.
- Betrachten das Modell $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, wobei ε ein Zufallsvektor der Dimension n mit $E(\varepsilon) = 0$ und $Cov(\varepsilon) = \sigma^2\mathbf{W}$. Dabei ist $\sigma^2 > 0$ und $\mathbf{W}$ eine symmetrische, positiv-definite Matrix.
- Es muss nun also weder Unabhängigkeit noch Varianzhomogenität gelten!
- Nehmen Normalverteilung an, d.h. $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{W})$ und $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{W})$.
- Nehme zudem an, dass $\text{Rang}(\mathbf{X}) = k$.

Beispiel: Gruppierte Daten

- m Gruppen $G_1, \dots, G_m$ mit $\mathbf{x}_i = \mathbf{x}_j = \mathbf{z}_l \in \mathbb{R}^k$ für alle i, j in Gruppe G_l
- z.B. wenn alle Kovariablen kategorial sind (Geschlecht, Kranheitsstadium, etc.)
- Daten können durch die Mittelwerte der Y_i einer Gruppe G_l zusammengefasst werden, d.h. durch

$$\bar{Y}_l = \frac{1}{n_l} \sum_{i \in G_l} Y_i, \qquad \text{wobei } n_l = |G_l|.$$

- Erhalten Daten der Form $(\bar{Y}_l, \mathbf{z}_l)$, $l = 1, \dots, m$, die stochastisch unabhängig sind.
- Wenn $Y_i \sim N(\mathbf{z}_l \beta, \sigma^2)$ für alle $i \in G_l$, dann $\bar{Y}_l \sim N(\mathbf{z}_l \beta, \frac{\sigma^2}{n_l})$, $l = 1, \dots, m$
- Also: $\bar{\mathbf{Y}} \sim N(\mathbf{Z}\beta, \sigma^2 \mathbf{W})$ mit $\mathbf{W} = \mathrm{diag}(1/n_1, ..., 1/n_m)$

Verhalten des klassischen KQ-Schätzers

- Betrachten den klassischen KQ-Schätzer $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$
- $\hat{\beta}$ ist immernoch erwartungstreu, d.h. es gilt $E(\hat{\beta}) = \beta$
- Für die Kovarianz gilt jetzt:

$$Cov(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$$

- Dies entspricht nicht mehr der Kovarianz aus dem klassischen linearen Modell. Die klassischen Test und Konfidenzintervalle würden also falsche Ergebnisse liefern.
- Werden sehen, dass $\hat{\beta}$ nun nicht mehr BLUE ist.

Der Aitken-Schätzer

- Betrachten nun transformierte Daten
$$\mathbf{Y}^* = \mathbf{W}^{-1/2}\mathbf{Y}, \quad \mathbf{X}^* = \mathbf{W}^{-1/2}\mathbf{X}, \quad \text{und } \varepsilon^* = \mathbf{W}^{-1/2}\varepsilon,$$
wobei $\mathbf{W}^{-1/2}$ eine Wurzel von $\mathbf{W}^{-1}$ ist (vgl. Vektor- und Matrixrechnung)

- Nun gilt
$$\mathbf{Y}^* = \mathbf{X}^*\beta + \varepsilon^*,$$
mit $E(\varepsilon^*) = 0$ und $Cov(\varepsilon^*) = \sigma^2\mathbf{I}$, d.h. $(Y_i^*, \mathbf{x}_i^*)$ erfüllen wieder die Voraussetzungen des klassischen Gauß-Markov-Modells.

Dies führt zum allgemeinen KQ-Schätzer

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

der auch *Aitken-Schätzer* genannt wird.

Eigenschaften des Aitken-Schätzers

- Man kann zeigen, dass

$$\hat{\beta}^{\mathbf{W}} = \text{argmin}_{\tilde{\beta} \in \mathbb{R}^k} (\mathbf{Y} - \mathbf{X}\tilde{\beta})^T \mathbf{W}^{-1} (\mathbf{Y} - \mathbf{X}\tilde{\beta})$$

- Im Falle einer Diagonalmatrix $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ ist

$$(\mathbf{Y} - \mathbf{X}\tilde{\beta})^T \mathbf{W}^{-1} (\mathbf{Y} - \mathbf{X}\tilde{\beta}) = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i \tilde{\beta})^2 / w_i$$

  die gewichtete Quadratsumme mit Gewichten $1/w_i$

- In diesem Fall wird der Aitken-Schätzer auch häufig *gewichteterer KQ-Schätzer* genannt

Eigenschaften des Aitken-Schätzers

Unter den Modell-Annahmen $E(\mathbf{Y}) = \mathbf{X}\beta$ und $Cov(\mathbf{Y}) = \sigma^2\mathbf{W}$ gilt:

- Erwartungswert

$$E(\hat{\beta}^{\mathbf{W}}) = (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}E(\mathbf{Y}) = (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}\beta = \beta,$$

  d.h. $\hat{\beta}^{\mathbf{W}}$ ist erwartungstreu,

- Kovarianz

$$\begin{aligned} Cov(\hat{\beta}^{\mathbf{W}}) &= (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}Cov(\mathbf{Y})\mathbf{W}^{-1^T}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1^T} \\ &= \sigma^2(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{W}\mathbf{W}^{-1^T}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1^T} \\ &= \sigma^2(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1} \end{aligned}$$

- Ist zudem $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{W})$, dann gilt $\hat{\beta}^{\mathbf{W}} \sim N(\beta, \sigma^2(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1})$

Eigenschaften des Aitken-Schätzers

Gauß-Markov-Theorem

Unter den Modell-Annahmen $E(\mathbf{Y}) = \mathbf{X}\beta$ und $Cov(\mathbf{Y}) = \sigma^2\mathbf{W}$ gilt, dass unter allen Schätzern $\tilde{\beta}$ von β, die linear und erwartungstreu sind, d.h. für die $\tilde{\beta} = \mathbf{AY}$ für eine $(k \times n)$-Matrix $\mathbf{A}$ und $E(\tilde{\beta}) = \beta$ gilt, der Aitken-Schätzer $\hat{\beta}^{\mathbf{W}}$ am besten ist, d.h. $Var(\mathbf{c}^T\hat{\beta}^{\mathbf{W}}) \leq Var(\mathbf{c}^T\tilde{\beta})$ für beliebige $\mathbf{c} \in \mathbb{R}^k$.
$\hat{\beta}^{\mathbf{W}}$ ist also BLUE.

Bemerkung: Das Theorem folgt aus dem klassischen Gauß-Markov-Theorem für $\mathbf{Y}^*, \mathbf{X}^*$

Prognosen und Residuen als Projektionen

Schiefwinklinge Projektion

- Aus dem Aitken-Schätzer ergeben sich die Prognosen und Residuen:

$$\hat{Y}_i = \mathbf{x}_i \hat{\beta}^{\mathbf{W}} \quad \text{und} \quad \hat{e}_i = Y_i - \mathbf{x}_i \hat{\beta}^{\mathbf{W}}$$

- Für geometrische Betrachtung brauchen wir die quadratische Form

$$(\mathbf{u}, \mathbf{v}) \longmapsto \mathbf{u}^T \mathbf{W}^{-1} \mathbf{v} =< \mathbf{u}, \mathbf{v} >_{\mathbf{W}}$$

- Alternativer Orthogonalitätsbegriff:

$$\mathbf{u} \perp_{\mathbf{W}} \mathbf{v} \quad \Longleftrightarrow \quad < \mathbf{u}, \mathbf{v} >_{\mathbf{W}} = \mathbf{u}^T \mathbf{W}^{-1} \mathbf{v} = 0 \quad \Longleftrightarrow \quad \mathbf{W}^{-1/2} \mathbf{u} \perp \mathbf{W}^{-1/2} \mathbf{v},$$

- Alternativer Längenbegriff

$$||\mathbf{u}||^2_{\mathbf{W}} =< \mathbf{u}, \mathbf{u} >_{\mathbf{W}} = \mathbf{u}^T \mathbf{W}^{-1} \mathbf{u}$$

Schiefwinklinge Projektion

- Auch für den alternativen Längenbegriff gilt der Satz des Pythagoras. Das heißt, gilt $\mathbf{a} \perp_{\mathbf{W}} \mathbf{b}$ für $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ so ist

$$\|\mathbf{a}+\mathbf{b}\|_{\mathbf{W}}^2 = \|\mathbf{a}\|_{\mathbf{W}}^2 + \|\mathbf{b}\|_{\mathbf{W}}^2$$

- $\mathbf{Y}$ lässt sich eindeutig als Summe eines Vektors $\mathbf{Y}^{\mathbf{W}} \in \mathcal{M}$ und eines Vektors $\hat{\mathbf{e}}^{\mathbf{W}} \in \mathcal{M}_{\mathbf{W}}^{\perp} = \{\tilde{\mathbf{e}} \in \mathbb{R}^n : < \tilde{\mathbf{e}}, \tilde{\mu} >_{\mathbf{W}} = 0 \text{ für alle } \tilde{\mu} \in \mathcal{M}\}$ zerlegen, d.h.

$$\mathbf{Y} = \hat{\mathbf{Y}}^{\mathbf{W}} \oplus_{\mathbf{W}} \hat{\mathbf{e}}^{\mathbf{W}}.$$

- Man kann zeigen, dass diese identisch zu den Prognosen und Residuen aus dem Aitken-Schätzer sind

Schiefwinklinge Projektion

- $\hat{\mathbf{Y}}$ und $\hat{\mathbf{e}}$ können nun wieder über eine Projektionsmatrix gewonnen werden:

$$\hat{\mathbf{Y}} = \mathbb{H}\mathbf{Y} \text{ mit der Projektionsmatrix } \mathbb{H} = \mathbf{P}_{\mathcal{M}}^{\mathbf{W}} = \mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}$$

- Es gilt wieder $\hat{\mathbf{e}} = \mathbf{Y} - \mathbb{H}\mathbf{Y} = (\mathbf{I} - \mathbb{H})\mathbf{Y} = (\mathbf{I} - \mathbb{H})\varepsilon$.
- Es ist leicht nachzuprüfen, dass $\mathbb{H}\mathbb{H} = \mathbb{H}$ und damit $(\mathbf{I} - \mathbb{H})\mathbb{H} = 0$ sowie $(\mathbf{I} - \mathbb{H})(\mathbf{I} - \mathbb{H}) = \mathbf{I} - \mathbb{H}$.
- Die Matrizen $\mathbb{H}$ und $\mathbf{I} - \mathbb{H}$ sind also beide Projektionsmatrizen; $\mathbb{H}$ projiziert auf $\mathcal{M}$ und $\mathbf{I} - \mathbb{H}$ auf $\mathcal{M}_{\mathbf{W}}^{\perp}$.
- Das heißt $\tilde{\mu} = \mathbb{H}\mathbf{Y}$ ist das Element in $\mathcal{M}$ mit kleinstem Abstand $\|\tilde{\mu} - \mathbf{Y}\|_{\mathbf{W}}^2$ zu $\mathbf{Y}$

Statistische Eigenschaften der Prognosen und Residuen

- $\hat{\mathbf{Y}}$ ist erwartungstreu:

$$E(\hat{\mathbf{Y}}) = E(\mathbb{H}\mathbf{Y}) = \mathbb{H}E(\mathbf{Y}) = \mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}\beta = \mathbf{X}\beta = \mu$$

- Für den Erwartungswert von $\hat{\mathbf{e}}$ gilt:

$$E(\hat{\mathbf{e}}) = E((\mathbf{I} - \mathbb{H})\mathbf{Y}) = E(\mathbf{Y} - \mathbb{H}\mathbf{Y}) = E(\mathbf{Y}) - E(\mathbb{H}\mathbf{Y}) = \mu - \mu = 0$$

- Für die Kovarianz von $\hat{\mathbf{Y}}$ gilt:

$$\begin{aligned} Cov(\hat{\mathbf{Y}}) &= \mathbb{H}Cov(\mathbf{Y})\mathbb{H}^T = \sigma^2\mathbb{H}\mathbf{W}\mathbb{H}^T \\ &= \sigma^2[\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}]\mathbf{W}[\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T] \\ &= \sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T \\ &= \sigma^2\mathbb{H}\mathbf{W} \end{aligned}$$

Statistische Eigenschaften der Prognosen und Residuen

- Ähnlich lässt sich zeigen, dass

$$Cov(\hat{\mathbf{e}}) = \sigma^2(\mathbf{W} - \mathbb{H}\mathbf{W}) = \sigma^2(\mathbf{I} - \mathbb{H})\mathbf{W}$$

- Für die Kovarianz zwischen $\hat{\mathbf{Y}}$ und $\hat{\mathbf{e}}$ lässt sich rechnen

$$Cov(\hat{\mathbf{e}}, \hat{\mathbf{Y}}) = \sigma^2(\mathbf{I} - \mathbb{H})\mathbf{W}\mathbb{H}^T = \sigma^2 \underbrace{(\mathbf{I} - \mathbb{H})\mathbb{H}}_{=\mathbf{0}} \mathbf{W} = \mathbf{0}.$$

- $\hat{\mathbf{e}}$ und $\hat{\mathbf{Y}}$ sind unkorreliert und damit bei normalverteiltem $\mathbf{Y}$ stochastisch unabhängig

Statistische Eigenschaften der Prognosen und Residuen

Ist $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{W})$ so folgt:

- Für die Verteilung von $\hat{\mathbf{Y}}$

$$\hat{\mathbf{Y}} \sim N\left(\mathbf{X}\beta, \sigma^2\mathbb{H}\mathbf{W}\right) = N\left(\mathbf{X}\beta, \sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\right)$$

- Für die Verteilung von $\hat{\mathbf{e}}$

$$\hat{\mathbf{e}} \sim N\left(\mathbf{0}, \sigma^2(\mathbf{I} - \mathbb{H})\mathbf{W}\right) = N\left(\mathbf{0}, \sigma^2(\mathbf{W} - \mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T)\right)$$

da $\mathbb{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}$

Schätzen der Residualvarianz

Schätzen der Residualvarianz

- Betrachte zunächst die sogenannte Devianz
$$Dev(\mathcal{M}) = ||\mathbf{Y} - \mathbf{X}\hat{\beta}||^2_{\mathbf{W}} = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{W}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \hat{\mathbf{e}}^T\mathbf{W}^{-1}\hat{\mathbf{e}}$$
- $Dev(\mathcal{M})$ misst den Abstand zwischen $\mathbf{Y}$ und den Prognosen $\hat{\mathbf{Y}}$ bezüglich des alternativen Abstandbegriffs
- Man kann zeigen, dass $Dev(\mathcal{M}) \sim \sigma^2\chi^2_{n-k}$
- Damit gilt $E(\hat{\mathbf{e}}^T\mathbf{W}^{-1}\hat{\mathbf{e}}) = \sigma^2(n-k)$
- Dies impliziert, dass
$$\hat{\sigma}^2 = \hat{\sigma}^2_{\mathbf{W}} = \frac{1}{n-k}\hat{\mathbf{e}}^T\mathbf{W}^{-1}\hat{\mathbf{e}}$$
ein erwartungstreuer Schätzer für σ^2 ist.

Der Aitken-Schätzer als MLE

- Ist $\mathbf{Y} \sim N(\mu, \sigma^2 \mathbf{W})$ so kann man zeigen, dass der Aitken-Schätzer $\hat{\beta}^{\mathbf{W}}$ auch der Maximum Likelihood Schätzer von β ist
- Außerdem kann man zeigen, dass der MLE für die Varianz gegeben ist durch:

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n} ||\mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathbf{W}}||^2_{\mathbf{W}} = \frac{1}{n} \hat{\mathbf{e}}^T \mathbf{W}^{-1} \hat{\mathbf{e}} = \frac{1}{n} ||\hat{\mathbf{e}}^{\mathbf{W}}||^2_{\mathbf{W}}.$$

- Dieser stimmt offensichtlich nicht mit dem erwartungstreuen Schätzer $\hat{\sigma}^2_{\mathbf{W}}$ überein

Testen linearer Hypothesen im allgemeinen linearen Modell

Lineare Hypothesen

- Betrachte wieder den Modellraum $\mathcal{M} = \{\mathbf{X}\tilde{\beta} \,|\, \tilde{\beta} \in \mathbb{R}^k\} \subseteq \mathbb{R}^n$
- Als Nullhypothese wird nun ein linearer Teilraum $\mathcal{M}_0 \subsetneq \mathcal{M}$ von $\mathcal{M}$ betrachtet
- $\mathcal{M}_0$ kann z.B. durch eine $(r \times k)$-dimensionale Restriktionsmatrix $\mathbf{C}$ definiert sein: $\mathcal{M}_0 = \{\mathbf{X}\tilde{\beta} \,|\, \tilde{\beta} \in \mathbb{R}^k \text{ mit } \mathbf{C}\beta = 0\}$
- Bezeichnen Dimension von $\mathcal{M}_0$ mit k_0
- Suchen einen statistischen Test für

$$H_0 : \mu \in \mathcal{M}_0 \quad \text{vs.} \quad H_1 : \mu \in \mathcal{M} \backslash \mathcal{M}_0$$

bzw. im Beispiel:

$$H_0 : \mathbf{C}\beta = 0 \quad \text{vs.} \quad H_1 : \mathbf{C}\beta \neq 0$$

Lineare Hypothesen - Beispiele

- Werden im Folgenden 2 Beispiele betrachten:
  (a) $\mathcal{M}_0 = \{\mathbf{X}\tilde{\beta} : \tilde{\beta}_1 \in \mathbb{R}, \tilde{\beta}_2 = \ldots = \tilde{\beta}_k = 0\}$. Das Modell $\mu \in \mathcal{M}_0$ entspricht dem Fall, in dem keine unabhängige Variable einen Einfluss auf Y hat.

  (b) Gegeben sei $S \subsetneqq \{2, \ldots, k\}$, z.B. die Koeffizienten aller Dummy-Variablen, die zu einem bestimmten Faktor gehören. Der Nullraum $\mathcal{M}_0 = \{\mathbf{X}\tilde{\beta} \,|\, \tilde{\beta} \in \mathbb{R}^k, \tilde{\beta}_j = 0$ für alle $j \in S\}$ entspricht der Situation, dass keine Dummy-Variable und damit der Faktor insgesamt keinen Einfluss auf Y hat.

Lineare Hypothesen - Prognosen und Teststatistik

- Betrachte zunächst die Prognosen:
  (1) für $\mathcal{M}$: $\hat{\mathbf{Y}} = \mathbf{P}^{\mathbf{W}}_{\mathcal{M}}\mathbf{Y}$ $(= \mathbb{H}^{\mathbf{W}}_{\mathcal{M}}\mathbf{Y})$,
  (2) für $\mathcal{M}_0$: $\hat{\mathbf{Y}}_0 = \mathbf{P}^{\mathbf{W}}_{\mathcal{M}_0}\mathbf{Y}$ $(= \mathbb{H}^{\mathbf{W}}_{\mathcal{M}_0}\mathbf{Y})$.

- Um einen Test für $H_0 : \mu \in \mathcal{M}_0$ gegen $H_1 : \mu \in \mathcal{M} \setminus \mathcal{M}_0$ zu definieren, betrachte nun die Zerlegung

$$\mathbf{Y} - \hat{\mathbf{Y}}_0 = \underbrace{\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0}_{\in \mathcal{M}} \oplus_{\mathbf{W}} \underbrace{\hat{\mathbf{e}}}_{\in \mathcal{M}^{\perp}_{\mathbf{W}}} .$$

- Damit folgt

$$||\mathbf{Y} - \hat{\mathbf{Y}}_0||^2_{\mathbf{W}} = ||\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0||^2_{\mathbf{W}} + ||\hat{\mathbf{e}}||^2_{\mathbf{W}} = (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0)^T \mathbf{W}^{-1} (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0) + \hat{\mathbf{e}}^T \mathbf{W}^{-1} \hat{\mathbf{e}}.$$

Lineare Hypothesen - Prognosen und Teststatistik

- Große Werte von $||\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0||^2_{\mathbf{W}}$ sprechen gegen $H_0 : \mu \in \mathcal{M}_0$ und für $H_1 : \mu \in \mathcal{M} \setminus \mathcal{M}_0$.
- Man kann zeigen, dass

$$\frac{1}{\sigma^2}(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0)^T \mathbf{W}^{-1}(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0) \underset{H_0}{\sim} \chi^2_{\dim(\mathcal{M})-\dim(\mathcal{M}_0)} = \chi^2_{k-k_0}$$

- Bei bekannter Residualvarianz σ^2 liefert die Verwerfungsregel

$$\frac{1}{\sigma^2}(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0)^T \mathbf{W}^{-1}(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0) \geq Q^{\chi^2}_{k-k_0}(1-\alpha)$$

also einen Niveau α Test für unser Testproblem.

Berechnung von $\hat{\mathbf{Y}}_0$ - Beispiel (a); Vereinfacht mit $\mathbf{W} =$diag$(w_1, ..., w_n)$

- $\mathcal{M}_0 = \{\mathbf{X}\tilde{\beta}|\tilde{\beta}_2, \ldots, \tilde{\beta}_n = 0\} = \{\tilde{\mu} = \mu_0 \mathbb{1}|\mu_0 \in \mathbb{R}\}$
- Unter $\mathcal{M}_0$ sind alle Erwartungswerte $E(\mathbf{Y}_i)$ gleich: $E(\mathbf{Y}_1) = \ldots = E(\mathbf{Y}_n)$.
- Der gewichtete KQ-Schätzer ist

$$\hat{\mu}_0 = \text{argmin}_{\mu_0 \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - \mu_0)^2 v_i \quad \text{mit} \quad v_i = 1/w_i.$$

- Man kann leicht zeigen, dass

$$\hat{\mu}_0 = \sum_{i=1}^{n} Y_i v_i / \sum_{i=1}^{n} v_i$$

  das gewichtete arithmetische Mittel ist

- Schließlich ist also $\hat{\mathbf{Y}}_0 = \hat{\mu}_0 \mathbb{1}$.

Berechnung von $\hat{\mathbf{Y}}_0$ - Beispiel (b)

- $\mathcal{M}_0 = \{\mathbf{X}\tilde{\beta} \,|\, \tilde{\beta}_j = 0 \text{ für } j \in S\}$, $S \subsetneq \{1, \dots, k\}$.
- Bilden die Matrix $\mathbf{X}_0$ durch Löschung der Spalten $\mathbf{x}_j$, $j \in S$, und definieren $s = |S|$
- Damit ist

$$\hat{\mathbf{Y}}_0 = \underbrace{\mathbf{X}_0}_{n\times(k-s)} \underbrace{(\mathbf{X}_0^T\mathbf{W}^{-1}\mathbf{X}_0)^{-1}}_{(k-s)\times(k-s)} \underbrace{\mathbf{X}_0^T\mathbf{W}^{-1}\mathbf{Y}}_{(k-s)\times n}$$

der Aitken-Schätzer ohne die Kovariablen in S.

F-Test für lineare Hypothesen

- Ist σ^2 unbekannt, so wird es geschätzt
- Wir wissen bereits, dass

$$(n-k)\frac{\hat{\sigma}^2_{\mathbf{W}}}{\sigma^2} = \frac{1}{\sigma^2}||\hat{\mathbf{e}}||^2_{\mathbf{W}} = \frac{1}{\sigma^2}\hat{\mathbf{e}}^T\mathbf{W}^{-1}\hat{\mathbf{e}} \sim \chi^2_{n-k}$$

- $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0$ und $\hat{\mathbf{e}}$ sind unkorreliert und damit unter der Annahme $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{W})$ stochastisch unabhängig
- Damit folgt

$$F = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0)^T\mathbf{W}^{-1}(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0)/(k-k_0)}{\hat{\mathbf{e}}^T\mathbf{W}^{-1}\hat{\mathbf{e}}/(n-k)} \sim F_{k-k_0,n-k},$$

wobei $F_{k-k_0,n-k}$ die F-Verteilung mit den Freiheitsgraden $df_1 = k - k_0$ und $df_2 = n - k$ ist

F-Test für lineare Hypothesen

- H_0 kann damit unter Kontrolle des Fehlers 1. Art auf dem Signifikanzniveau α verworfen werden mit der Verwerfungsregel

$$F \geq Q^F_{k-k_0,n-k}(1-\alpha).$$

- Die Entscheidung kann äquivalent über einen p-Wert getroffen werden:

$$p = 1 - F^F_{k-k_0,n-k}(F) \leq \alpha \ .$$

Alternative Darstellung der F-Statistik für lineare Hypothesen

- Es gilt
$$||\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0||^2_{\mathbf{W}} = ||\hat{\mathbf{e}}_0||^2_{\mathbf{W}} - ||\hat{\mathbf{e}}||^2_{\mathbf{W}} = \hat{\mathbf{e}}_0^T \mathbf{W}^{-1} \hat{\mathbf{e}}_0 - \hat{\mathbf{e}}^T \mathbf{W}^{-1} \hat{\mathbf{e}}$$
wobei $\hat{\mathbf{e}}_0 = \mathbf{Y} - \hat{\mathbf{Y}}_0$ das Residuum des Nullmodels $\mathcal{M}_0$ ist

- Wir erhalten somit
$$F = \frac{(\hat{\mathbf{e}}_0^T \mathbf{W}^{-1} \hat{\mathbf{e}}_0 - \hat{\mathbf{e}}^T \mathbf{W}^{-1} \hat{\mathbf{e}})/(k - k_0)}{\hat{\mathbf{e}}^T \mathbf{W}^{-1} \hat{\mathbf{e}}/(n - k)}.$$

- Man nennt den allgemeinen F-Test auch *Devianz-Analyse*, da $\hat{\mathbf{e}}_0 \mathbf{W}^{-1} \hat{\mathbf{e}}_0 = ||\hat{\mathbf{e}}_0||^2 = ||\mathbf{Y} - \hat{\mathbf{Y}}_0^{\mathbf{W}}||^2$ und $\hat{\mathbf{e}}^T \mathbf{W}^{-1} \hat{\mathbf{e}} = ||\hat{\mathbf{e}}||^2 = ||\mathbf{Y} - \hat{\mathbf{Y}}^{\mathbf{W}}||^2$ die *Devianzen* von $\mathcal{M}$ und $\mathcal{M}_0$ sind

Modelle mit autokorrelierten Fehlern

Allgemeines Setup

- Beobachtungen $Y_1, Y_2, \dots, Y_n$ seien nun zeitlich geordnete Messungen einer einzigen Beobachtungseinheit
- Da sie der selben Beobachtungseinheit entstammen, sind $Y_1, Y_2, \dots, Y_n$ in der Regel korreliert
- Wenn $Y_i = \mathbf{x}_i\beta + \epsilon_i$, dann ist davon auszugehen, dass die Störterme $\epsilon_1, \dots, \epsilon_n$ korreliert sind
- Werden im Folgenden verschiedene Modelle für die Kovarianzstruktur der ϵ_i betrachten

Das AR(1)-Modell für die Störterme

- Im AR(1)-Modell nimmt man an, dass für $\rho \in (-1, 1)$ und $\sigma^2 > 0$

$$Cov(\epsilon) = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

- Korrelation kann auch durch die sogen. *Autokorrelationsfunktion* beschrieben werden:

Autokorrelationsfunktion

$$\mathrm{ACF}(j) = Cov(\epsilon_i, \epsilon_{i-j}) = \frac{Cov(\epsilon_i, \epsilon_{i-j})}{Var(\epsilon_i)}$$

Autokorrelationsfunktion

Autokorrelationsfunktion

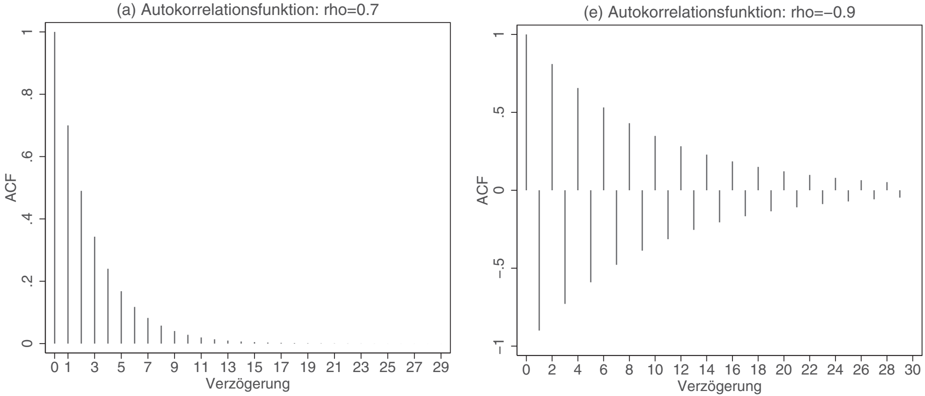
$$\text{ACF}(j) = Cov(\epsilon_i, \epsilon_{i-j}) = \frac{Cov(\epsilon_i, \epsilon_{i-j})}{Var(\epsilon_i)}$$

- Die Autokorrelationsfunktion beschreibt die Korrelation zweier Beobachtungen Y_i und Y_{i-j} in Abhängigkeit von ihrem zeitlichen Abstand j
- Im AR(1)-Modell fällt sie mit dem zeitlichen Abstand j ab und es gilt:

$$\text{ACF}(j) = \rho^j$$

- Falls $\rho > 0$ fällt die Korrelation geometrisch, für $\rho < 0$ hat man einen alternierenden Abfall (siehe nächste Folie)

Autokorrelationsfunktion im AR(1) Modell



Autokorrelationfunktion bei AR(1)-Korrelation für $\rho = 0.7$ bzw. $\rho = -0.9$; aus Fahrmeir, Kneib und Lang (2009)

Stochastisches Modell des AR(1) Modells

- Die Kovarianzstruktur des AR(1)-Modells kann über unendliche Folge von stochastisch unabhängigen Zufallsvariablen erzeugt werden

- Es seien

$$U_n, U_{n-1}, U_{n-2}, \ldots, U_1, U_0, U_{-1}, U_{-2}, \ldots$$

  stochastisch unabhängig mit $E(U_i) = 0$ und $Var(U_i) = \sigma^2$ für alle i.

- Die AR(1)-korrelierten Störterme ϵ_i können dann durch die folgenden unendlichen Summen erzeugt werden:

$$\epsilon_i = \sum_{k=0}^{\infty} \rho^k U_{i-k}$$

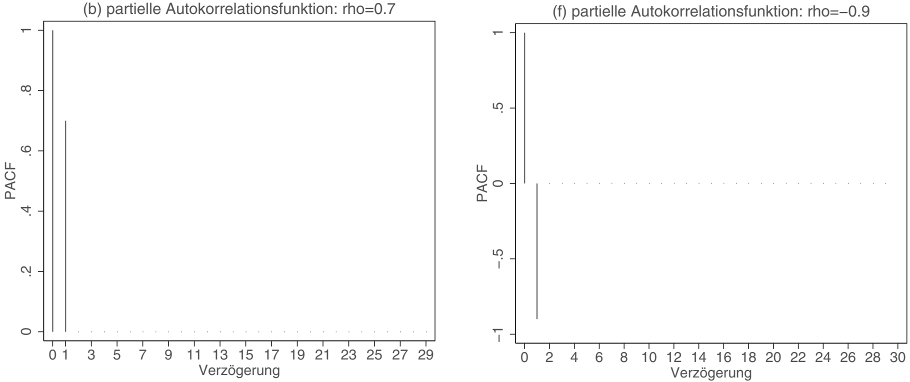
Stochastisches Modell des AR(1) Modells

- Man kann zeigen, dass dann
$$\epsilon_i = \rho\epsilon_{i-1} + U_i$$
- $\epsilon_{i-1} = \sum_{k=0}^{\infty} \rho^k U_{i-1-k}$ und U_i sind stoch. unabh.
- ϵ_i erfüllt das lineare Modell mit ϵ_{i-1} als unabhängige Variable
- Sprechen daher von *Autokorrelation der Ordnung 1*, abgekürzt *AR(1)*
- Ordnung 1, da ϵ_i nur vom letzten ϵ_j abhängt

Autokorrelation höherer Ordnung

- Man spricht von Autokorrelation der Ordnung j für ein $j > 1$ und schreibt AR(j), wenn
$$\epsilon_i = \alpha_1 \epsilon_{i-1} + \alpha_2 \epsilon_{i-2} + \ldots + \alpha_j \epsilon_{i-j} + U_i,$$
mit $E(U_i) = 0$, $Var(U_i) = \sigma^2$, stochastisch unabhängig von $\epsilon_1, \ldots, \epsilon_{i-1}$
- Man betrachtet die sogenannte partielle Autokorrelationsfunktion:
$$j \mapsto \mathrm{PACF}(j) = \alpha_j$$
- Liefert Regressionskoeffizienten von ϵ_{i-j} des lin. Modells für ϵ_i mit Kovariablen $\epsilon_{i-1}, \ldots, \epsilon_{i-j}$.

PACF im AR(1) Modell



PACF des AR(1)-Modells mit $\rho = 0.7$ bzw. $\rho = -0.9$; aus Fahrmeier, Kneib und Lang (2009)

Diagnose in Daten

Empirische ACF und PACF

- Um anhand von Daten zu prüfen werden die empirische Autokorrelationsfuktion

$$\widehat{ACF}(j) = \frac{\widehat{Cov}(\epsilon_i, \epsilon_{i-j})}{\widehat{Var}(\epsilon_i)},$$

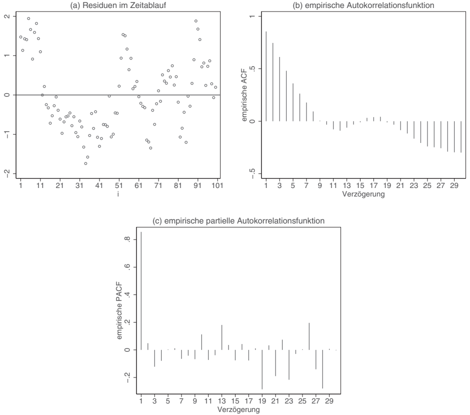
wobei $\widehat{Cov}(\epsilon_i, \epsilon_{i-j}) = \frac{1}{n}\sum_{i=j+1}^{n} \hat{\epsilon}_i \hat{\epsilon}_{i-j}$,

- und die empirische partielle Autokorrleationsfunktion

$$\widehat{PACF}(j) = \hat{\alpha}_j,$$

wobei $\hat{\alpha}_j$ KQ-Schätzer einer lin. Regression von $\hat{\epsilon}_i$ nach $\hat{\epsilon}_{i-1}, \ldots, \hat{\epsilon}_{i-j}$ ist, betrachtet

Beispiel: AR(1) Modell



Aus Fahrmeier, Kneib und Lang (2009)

Schätzung der Parameter

Setup

- Bisher Annahme, dass **W** bekannt ist und nur σ^2 und β geschätzt werden müssen
- Nun auch **W** unbekannt
- Benötigen Annahmen um die Zahl der Parameter bei der Schätzung von **W** zu reduzieren
- Hier: Schätzmethoden im AR(1)-Modell (dann muss nur ρ geschätzt werden um **W** zu bestimmen)

Zweistufige Schätzung

1. Schätzer $\hat{\beta}$ als (klassischen) KQ-Schätzer und bestimme zugehörige Residuen $\hat{\mathbf{e}}$

2. Schätze die Autokorrelation ρ

$$\hat{\rho} = \frac{\sum_{i=2}^{n} \hat{e}_i \hat{e}_{i-1}}{\sqrt{\sum_{i=2}^{n} \hat{e}_i} \sqrt{\sum_{i=2}^{n} \hat{e}_{i-1}}}.$$

3. Bilde die Gewichtsmatrix

$$\hat{\mathbf{W}} = \frac{1}{1-\hat{\rho}^2} \begin{pmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \dots & \hat{\rho}^{n-1} \\ \hat{\rho} & 1 & \hat{\rho} \dots & \hat{\rho}^{n-2} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{\rho}^{n-1} & \hat{\rho}^{n-2} & \hat{\rho}^{n-3} & \dots & 1 \end{pmatrix},$$

und schätze β mit dem Aitken-Schätzer mit $\mathbf{W} = \hat{\mathbf{W}}$.

Zweistufige Schätzung

- Schritte 1.-3. können zur Verbesserung der Anpassung iteriert werden
- Man nennt dies die *Prais-Winston-Schätzung*
- Unter recht allgemeinen Bedingungen: die Schätzung ist konsistent

Maximum-Likelihood-Methode

- Unter der Normalverteilungsannahme für $\mathbf{Y}$ ist die log-Likelihood der Daten gegeben als:

$$l(\beta, \sigma, \rho) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \frac{1}{2}(1-\rho^2) - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{W}^{-1}(\mathbf{Y} - \mathbf{X}\beta)/\sigma^2$$

- Maximiere durch nullsetzen des Scores (Ableitung der log-Likelihood)
- Numerische Lösung des resultierenden Gleichungssystems