

Auf diesem Aufgabenblatt wiederholen und vertiefen wir Modelldiagnose und Modellwahl im linearen gemischten Modell (Vorlesungsfolien 7) und beginnen mit der Betrachtung nicht normalverteilter longitudinaler Daten und ihrer Modellierungsmöglichkeiten (Vorlesungsfolien 8).

Aufgabe 1: Modelldiagnose

In dieser Aufgabe beschäftigen wir uns mit der Diagnose von geschätzten linearen gemischten Modellen, um die getroffenen Annahmen und Spezifikationen zu überprüfen.

- (a) Laden Sie den Datensatz `vitamin` von der Homepage herunter und lesen Sie die Beschreibung. Verschaffen Sie sich einen ersten Überblick.
- (b) Schätzen Sie ein lineares gemischtes Modell (`m_RIRS`) mit zufälligen Interzepts und zufälligen Steigungen für jedes Kind und festen Effekten für `time` und für die Interaktion von `group` und `time`, aber ohne Haupteffekt für `group`. Nehmen Sie an, dass keine serielle Korrelation vorliegt.
- (c) Wieso braucht man hier keinen Haupteffekt für `group` zu betrachten?
- (d) Sie möchten nun den Modellfit bewerten. Plotten sie die populations-spezifischen Residuen $r_{ij} = y_{ij} - x_{ij}^\top \hat{\beta}$ gegen die jeweiligen Mittelwertsschätzer $\hat{y}_{ij} = x_{ij}^\top \hat{\beta}$. Welche beiden Modellschwächen lassen prinzipiell sich dadurch aufzeigen? Interpretieren Sie den Plot.
Hinweis: Sie können das Argument `level` in der Funktion `resid(m_RIRS)` verwenden, um die Residualebene zu bestimmen.
- (e) Welches Problem ergibt sich bei der Erstellung eines Quantil-Quantil Plots für die Residuen r_{ij} ? Wie können sie geeignet transformiert werden?

Aufgabe 2: Modellwahl

In dieser Aufgabe geht es darum verschiedene lineare gemischten Modellen zu vergleichen und das passendere Modell auszuwählen.

- (a) In Aufgabe 1 haben wir gesehen, dass der Mittelwert nicht gut spezifiziert wurde, daher kann es Sinn machen, anstelle von `time` die transformierte Variable `log(time)` zu verwenden. Schätzen Sie im Folgenden ein Modell `m_RIRSlog`, das mit dem Modell aus Aufgabe 1 bis auf die Transformation von `time` identisch ist. Verwenden Sie zur Schätzung **ML** (statt REML) und schätzen Sie auch das Modell `m_RIRS` noch einmal **mit ML**.
Hinweis: Beachten Sie, dass auch die zufälligen Steigungen mit der Transformation angepasst werden.
- (b) Wie lassen sich Modell `m_RIRS` und Modell `m_RIRSlog` vergleichen, d.h. wie kann man

auswählen welches der beiden Modelle geeigneter ist? Zu welcher Entscheidung kommen Sie bei ihrer Modellselektion?

- (c) Welches Problem würde auftreten, wenn die beiden obigen Modelle, die wir vergleichen wollen, mit REML geschätzt würden?
- (d) Welcher Blickwinkel auf das Modell wird in der Betrachtung des marginalen bzw. des konditionalen AIC verkörpert?
- (e) Vergleichen Sie im Folgenden das Modell `m.RIRS1og` und dasselbe Modell aber ohne zufällige Steigungen. Welches Modell würde auf Basis des marginalen AIC gewählt werden?
- (f) Verwenden Sie stattdessen das konditionale AIC (aus R-package `cAIC4`) um die beiden Modelle zu vergleichen. Zu welcher Entscheidung kommen Sie nun?

Hinweis: Um die Funktion `cAIC4::cAIC` verwenden zu können, schätzen Sie die Modelle mit der Funktion `lmer` aus Paket `lme4` (vgl. Vorlesungsfolien). Betrachten Sie die Hilfe `?lme4`, um die wesentlichen Unterschiede zur bereits bekannten Funktion `lme` zu verstehen.

Aufgabe 3: Nicht normalverteilte longitudinale Daten

Im Datensatz `epil` im R-Paket `MASS` liegen die Anzahlen an epileptischen Anfällen für 59 Epileptiker über die Zeit vor. Die Probanden wurden einer Behandlungsgruppe mit dem Medikament Progabide und einer Placebogruppe randomisiert zugeteilt.

- (a) Welches Modell würden Sie wählen, wenn sie sich für individuelle Prognosen für die Probanden interessieren?
- (b) Welches Modell würden Sie wählen, wenn sie sich für den Einfluss des Medikaments Progabide auf Populationsniveau interessieren?
- (c) Sei $\hat{\beta}$ der geschätzte Einfluss von Progabide aus dem Modell in (a). Was müssten Sie bei der Interpretation von $\hat{\beta}$ berücksichtigen? Entspricht die Interpretation der des Effekts von Progabide aus dem Modell in (b)?