Analyse Longitudinaler Daten SoSe 2017 **Blatt 4**
Prof. Dr. Sonja Greven und Almond Stöcker 20.06.2017

Auf diesem Aufgabenblatt beschäftigen wir uns mit dem Testen von zufälligen Effekten, sowie mit flexiblen Erweiterungen des linearen gemischten Modells (LMM). Die zu bearbeitenden Aufgaben beziehen sich auf die Inhalte der fünften und sechsten Vorlesungsfolien.

Aufgabe 1: Testen von zufälligen Effekten

Im Folgenden soll die Notwendigkeit ausgewählter zufälliger Effekte anhand des Datensatzes antibiotics getestet werden. Der Datensatz (im `groupedData` Format) und eine kurze Beschreibung sind auf der Homepage verfügbar.

a) Auf was wird getestet, wenn auf die Notwendigkeit eines zufälligen Effekts getestet wird?

b) Schätzen Sie ein Modell `m_RI` mit zufälligen Interzepts für jedes Kind und ein Modell `m_RIRS` mit zufälligen Interzepts und zufälligen Steigungen für jedes Kind mit jeweils `time` als festem Effekt. *Hinweis:* Verwenden Sie die Funktion `update()`.

c) Führen Sie mit der Funktion `anova()` (`?anova.lme()`) einen Likelihood-Quotienten (LQ) Test durch, um auf die Notwendigkeit der subjektspezifischen Trends zu testen.

   (i) Formulieren Sie die zugehörige Nullhypothese.

   (ii) Wie lautet die Testentscheidung?

   (iii) Warum ist die Verwendung dieses Tests, so wie er in `R` implementiert ist, problematisch?

d) Häufig wird die Mischung zweier χ^2 - Verteilungen als approximative Verteilung der LQ-Statistik unter H_0 verwendet. Berechnen Sie auf diese Weise den sich ergebenden p-Wert und vergleichen Sie diesen mit dem Ergebnis aus **c)**.

e) Ist es bei diesem Test problematisch, dass die Modelle mit REML geschätzt wurden? Begründen Sie Ihre Antwort.

f) Schätzen Sie nun ein Modell `m_RIRS_unc`, das sich von `m_RIRS` lediglich darin unterscheidet, dass zufällige Interzepts und zufällige Steigungen als unkorreliert angenommen werden.

   (i) Formulieren Sie die Nullhypothese.

   (ii) Führen Sie anschließend einen Likelihood-Quotienten Test durch, der überprüft, ob der zufällige Interzept und die zufällige Steigung korreliert sind.

   (iii) Ist die gängige Verteilungsannahme der LQ-Statistik in diesem Fall gültig?

Aufgabe 2: Flexible Erweiterungen für den Mittelwert

In dieser Aufgabe beschäftigen wir uns mit flexiblen Annahmen über den Zusammenhang zwischen Kovariablen und Response, der nicht ausschließlich linear sein muss.

a) Lesen Sie sich die `R`-Hilfe zu Funktion `gamm` in Paket `mgcv` durch.

b) Betrachten Sie nun erneut den Datensatz antibiotics und schätzen Sie ein LMM mit unkorrelierten zufälligen Interzepts und zufälligen Steigungen für jedes Kind, sowie einer glatten Funktion von `time`.

c) Stellen Sie den Effekt von `time` graphisch dar und interpretieren Sie ihn.
Hinweis: Verwenden Sie hierfür die Funktion `plot.gam`.

d) Schreiben Sie die geschätzte Kovarianzmatrix der zufälligen Effekte hin.
Hinweis: Verwenden Sie hierzu die `summary`-Funktion.

Aufgabe 3: Flexible Erweiterungen für die Fehlervarianz

In dieser Aufgabe beschäftigen wir uns mit Annahmen für die Fehlervarianz $\mathrm{Cov}(\epsilon_i) = \boldsymbol{\Sigma}_i$ im LMM. Wir verwenden hierzu den bereits bekannten Datensatz rats, der ebenfalls auf der Homepage verlinkt ist.

a) Was impliziert die häufig getroffene Annahme $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i}$? Was hat das für Konsequenzen in einem Modell mit zufälligen Interzepts?

b) In welcher formalen Form lässt sich serielle Korrelation im `R`-Paket `nlme` (mittels `corStruct`, vgl. letzte Übungen) spezifizieren?

c) Wie lässt sich die Kovarianz bei serieller Korrelation im LMM weiter zerlegen? Schreiben Sie die betreffenden Modellkomponenten, sowie die zugehörigen Annahmen hin.

d) Geben Sie (ohne Berechnung) die marginale Korrelation zwischen zwei Messungen an einem Subjekt für ein Modell mit zufälligen Intercepts und serieller Korrelation an und erklären Sie die Bestandteile.

e) Eine Erweiterung durch serielle Korrelation kann zwar sinnvoll sein, bedeutet allerdings auch, dass mehr Parameter geschätzt werden müssen. Daher bietet es sich an zu überprüfen, ob serielle Korrelation wirklich notwendig ist.

   (i) Wie lässt sich serielle Korrelation beurteilen? Skizzieren Sie kurz das Vorgehen.

   (ii) Verwenden Sie den auf der Homepage zur Verfügung gestellten Code um das folgende lineare Modell für den Datensatz rats zu schätzen

$$\texttt{REPONSE}_{ij} = \beta_0 + \beta_1 \texttt{logT}_{ij} + \beta_2 \texttt{GROUP1}_i \cdot \texttt{logT}_{ij} + \beta_3 \texttt{GROUP2}_i \cdot \texttt{logT}_{ij} + \varepsilon_{ij}.$$

   Bestimmen Sie das empirisches Semi-Variogramm und plotten Sie dieses.
   Hinweis: Code für die Bestimmung eines empirischen Semi-Variogramms finden Sie auf der Homepage.

   (iii) Was können Sie dem Plot entnehmen?