

Auf diesem Aufgabenblatt beschäftigen wir uns mit generalisierten linearen gemischten Modellen (GLMMs). Die zu bearbeitenden Aufgaben beziehen sich auf die Inhalte der achten und neunten Vorlesungsfolien.

### Aufgabe 1: GLMM für binäre Daten

Der Datensatz `madras` im Paket `drm` umfasst die monatlichen Aufzeichnungen (Vorhandensein/Ausbleiben) des psychiatrischen Symptoms “thought disorder” (Verwirrtheit) von 86 Patienten während des ersten Jahres ab Ersteinlieferung ins Krankenhaus.

- (1) Laden Sie den Datensatz und verschaffen Sie sich einen ersten Überblick über die Daten.
- (2) Wie viele Beobachtungen liegen für die Patienten jeweils vor? Handelt es sich um einen balancierten Datensatz?
- (3) Wandeln Sie die Variable `id` in eine Faktorvariable mit Levels 1 bis 86 um.
- (4) Berechnen Sie nun für die Frauen und Männer getrennt jeweils die relativen Häufigkeiten des Auftretens von “thought disorder” im zeitlichen Verlauf und plotten Sie die Verläufe. Was lässt sich dem Plot entnehmen?
- (5) Da man sich bei diesen Daten für die individuen-spezifischen Effekte interessiert, soll ein GLMM mit festen Effekten für `month` und `gender`, sowie zufälligen Interzepts pro Patient geschätzt werden. Formulieren Sie das Modell und spezifizieren Sie die Verteilungsannahmen, die Strukturannahme (systematische Komponente), sowie eine geeignete Linkfunktion.
- (6) Was unterscheidet die Schätzung eines GLMM von der eines LMM? Was macht die Schätzung so aufwändig?
- (7) Welche drei Möglichkeiten zur Schätzung von GLMMs haben Sie kennengelernt, was wird dabei jeweils approximiert?
- (8) Schätzen Sie das Modell mit jeder dieser drei Varianten. Verwenden Sie hierfür die in der Vorlesung vorgestellten Funktionen aus den Paketen `MASS` und `lme4`. Verwenden Sie für die adaptive Gauß-Quadratur (AGQ) 11 Quadraturpunkte. Vergleichen Sie die Schätzungen der festen Effekte.
- (9) Weshalb dauert die Schätzung mit AGQ länger als die Laplace-Schätzung?
- (10) Was sollten Sie bei der Interpretation der Ergebnisse der PQL-Schätzung bei diesem Datensatz beachten?
- (11) Interpretieren Sie die Schätzer der festen Effekte der AGQ-Schätzung.

- (a) Wieso kann man die Koeffizientenschätzer i.A. nicht marginal interpretieren?
  - (b) Plotten Sie die individuellen Wahrscheinlichkeiten  $P(Y_{ij} = 1|b_i)$  getrennt nach Geschlecht über die Zeit (`month`).
  - (c) Fügen Sie den beiden Plots die Wahrscheinlichkeitskurve über die Zeit für  $b_i = 0$  hinzu.
  - (d) Berechnen Sie nun die mittleren Wahrscheinlichkeiten  $P(Y_{ij} = 1)$  für Männer und Frauen getrennt und fügen Sie auch diese den Plots hinzu. Wie lassen sich die resultierenden Plots interpretieren?
- (12) Sie möchten nun mit einem LQ-Test testen, ob es eine Interaktion zwischen dem Geschlecht der Patienten und der Zeit gibt.
- (a) Wieso eignet sich eine Schätzung basierend auf PQL dafür nicht?
  - (b) Wieso gibt die `summary`-Funktion in R für die Schätzung mittels PQL für das AIC keinen Wert (`NA`) aus?
- (13) Verwenden Sie stattdessen die AGQ-Schätzung und führen Sie den in (12) vorgeschlagenen Test durch. Wie lautet Ihre Testentscheidung?
- (14) Ein alternatives Vorgehen um longitudinale nicht normalverteilte Daten zu analysieren ist die Verwendung marginaler Modelle (nächstes Blatt). Worin liegen die Hauptunterschiede zwischen diesen beiden Ansätzen?