

Aufgabe 1

Laden Sie sich den Datensatz *Amsterdam* von der Homepage der Vorlesung herunter.

Beschreibung des Datensatzes:

- Response *TC*: Gesamt-Serumcholesterol (mmol/l) von 147 Probanden insgesamt 6 Mal pro Proband gemessen
 - Kovariablen:
 - *t*: Zeit: Beginn der Studie 1977 ($t = 0$) alle Probanden 13 Jahre alt; ersten vier Messungen im Jahresabstand, letzten beiden Messungen 1985 und 1991
 - *bfatness*: Körperfettmessung (zeitvariierend): Summe der Dicke von 4 Hautfalten (mm)
 - *smoker*: Raucher (ja/nein; zeitvariierend)
 - *gender*: Geschlecht
 - *fitness*: Fitness-Level bei Aufnahme in die Studie (also bei $t=0$). Gemessen als maximale Sauerstoff-Aufnahme bei einem Laufbandtest.
- (a) Erstellen Sie mit der Funktion `xypplot()` einen Spaghettiplot der TC-Messungen für die vier möglichen Kombinationen von *smoker* und *gender*.
- (b) Passen Sie an die Daten ein LMM mit Intercept, zeitlichem Trend sowie Random Intercept und Random Slope an. Geben Sie die Modellformel und die entsprechenden Verteilungsannahmen für die vorkommenden Zufallsvariablen an. Identifizieren Sie aus dem Output die Schätzer aller auftretenden Parameter.
- (c) Bestimmen Sie die marginale Kovarianzmatrix der Beobachtungen eines Individuums aus der Studie und interpretieren Sie diese. Wie sieht die marginale Kovarianzmatrix des gesamten Responsevektors aus?
- (d) *Zusatzaufgabe*: Skizzieren Sie mit Hilfe eines Plots von *t* gegen *y*, was positive bzw. negative Korrelation zwischen zufälligem Intercept und zufälliger Steigung im LMM bedeuten.
- (e) In R wird folgendes Modell spezifiziert:

```
library(mgcv)
mod <- gamm(tc ~ 1 + s(t, by=gender, k=5) + bfatness + smoker,
            random= list(id=~1+t), data=Amsterdam)
```

Geben Sie die Modellgleichung an. Vergessen Sie dabei nicht, Verteilungsannahmen für die zufälligen Komponenten anzugeben. Schätzen Sie das Modell in R und erstellen Sie Graphiken der glatten Effekte.

Aufgabe 2

In einer Studie soll die Wirksamkeit eines Medikaments gegen Epilepsie untersucht werden. Dabei erhalten die Patienten ein Placebo (medikament = 0) oder ein Medikament (medikament = 1). Als weitere Kovariate wurde das Alter der Patienten in Jahren erhoben (alter). Die interessierende Zielvariable y ist die Anzahl der Anfälle pro Woche. Es wurden insgesamt 65 Patienten ($id=1, \dots, 65$) in 7 Studienzentren ($linik=1, \dots, 7$) über einen Zeitraum von jeweils 5 Wochen untersucht.

Ein Statistiker passt folgendes generalisiertes, lineares gemischtes Modell auf den Daten an:

```
> library(MASS)
> glmm <- glmmPQL(y ~ medikament + alter, random = ~ 1|linik/id,
+               family = poisson(link = "log"), data = data, verbose = FALSE)
> glmm
[...]
Fixed: y ~ medikament + alter
(Intercept)  medikament      alter
0.05662618  0.35774136  0.01453602

Random effects:
Formula: ~1 | klinik
      (Intercept)
StdDev:    0.157035

Formula: ~1 | id %in% klinik
      (Intercept)  Residual
StdDev:    0.4600308 0.9518473
[...]

> ranef(glmm)
Level: klinik                Level: id %in% klinik
  (Intercept)                (Intercept)
1 -0.06574077                1/3   0.53322220
2 -0.04165915                1/10 -0.04494309
3 -0.03703619                1/15 -0.62865473
[...]                        [...]
```

- Geben Sie die Modellgleichung des geschätzten generalisierten linearen gemischten Modells an. Benennen / beschreiben Sie alle Modellkomponenten (Modellgleichung, Response, feste Effekte, zufällige Effekte, Indizes) sowie die getroffenen Verteilungsannahmen.
- Interpretieren Sie den Parameter zur Gabe des Medikaments.
- Berechnen Sie die erwartete Anzahl von Anfällen pro Woche für den Patienten mit $id=3$. Dieser Patient wurde in Klinik 1 behandelt, erhielt das Medikament und ist 52 Jahre alt.
- Das Modell wurde mit Penalized Quasi-Likelihood (PQL) angepasst. Erklären Sie kurz die Prinzipien hinter diesem Schätzalgorithmus. Welche Likelihood wird dabei maximiert? Geben Sie ein Argument, warum der Algorithmus für die vorliegenden Daten geeignet bzw. nicht geeignet ist.