

Aufgabe 1

Laden Sie sich den Datensatz *Amsterdam*, eine Wachstums- und Gesundheitsstudie, von der Homepage der Vorlesung herunter.

Kurze Beschreibung des Datensatzes:

- Gesamt-Serumcholesterol 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.
 - Response: TC: Gesamt-Serumcholesterol (mmol/l)
- (a) Erstellen Sie mit der Funktion `xyp1ot()` einen Spaghettiplot der TC-Messungen für die vier möglichen Kombinationen von *smoker* und *gender*
- (b) Passen Sie an die Daten ein einfaches LMM mit 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 den beiden Random-Effects im einfachen LMM bedeutet.
- (e) In R wird folgendes Modell spezifiziert:

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

Geben Sie die Modellgleichung an. Vergessen Sie dabei nicht Verteilungsannahmen für die zufälligen Komponenten anzugeben. Geben Sie außerdem an, wie ein glatter Effekt mit Hilfe von festen und zufälligen Effekten dargestellt werden kann. Verwenden Sie für diese Darstellung die TP-Basis.

Schätzen Sie das Modell in R und erstellen Sie Graphiken der glatten Effekte.

- (f) Zusatzaufgabe: Schätzen Sie für jede Person ein individuelles Modell. (Dazu kann man die Funktion `lmList()` aus dem Paket `nlme` verwenden auf ein `groupedData`-Objekt anwenden.) Vergleichen Sie anschließend die Koeffizienten aus dem gemischten Modell von Teilaufgabe (b) mit den Koeffizienten aus den individuellen Modellen.

Aufgabe 2

Nehmen Sie an, dass die interessierende Zielgröße ist, ob ein Patient in eine kardiovaskuläre Risiko-
gruppe fällt oder nicht. Für jeden Zeitpunkt t ist der kardiovaskuläre Risikoindikator Y_{it}^* (0=klein,
1=groß) von Individuum i gegeben durch

$$Y_{it}^* I(Y_{it} > q_{0.75}(t))$$

mit $q_{0.75}(t)$ 75%-Quantil der beobachteten TC-Messungen zum Zeitpunkt t .

- (a) Fitten Sie ein gemischtes logistisches Regressionsmodell mit festen Effekten (nur Haupteffekte) für t , *gender* und *smoker* sowie einem Random-Intercept mit AGQ.
- (b) Geben Sie die entsprechende Modellformel für dieses Modell an und identifizieren Sie aus dem Output die Schätzer für alle relevanten Größen. Interpretieren Sie den Regressionskoeffizienten für *smoker* auf der Odds-Ratio-Skala.