Aufgabe 1:

In dieser Aufgabe sollen Sie den FEV-Datensatz analysieren. Dieser enthält wiederholte Messungen eines Lungenleistungsparamters (FEV= Forciertes Exspirations-Volumen) an 300 Mädchen im Alter zwischen 6 bis 19 Jahren. Benutzen Sie den folgenden Code um den FEV-Datensatz von der Homepage zu laden und die Daten zu visualisieren:

```
R>library(nlme)
R>url <- "http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/longitudinal_ss08/"
R>fev <- groupedData(log.fev1 ~ age | subject, read.table(paste(url,
+     "download/fev1.txt", sep = ""), header = T))
R>groups <- unique(fev$subject)
R>xl <- range(fev$age)
R>yl <- range(fev$log.fev1)
R>for (i in 1:30) {
+     samp <- groups[(1 + (i - 1) * 10):(i * 10)]
+     d <- subset(fev, subject %in% samp)
+     print(xyplot(log.fev1 ~ age, data = d, groups = subject,
+         type = "b", xlim = xl, ylim = yl))
+     Sys.sleep(0.5)
+ }
R>xl <- range(fev$height)
R>for (i in 1:30) {
+     samp <- groups[(1 + (i - 1) * 10):(i * 10)]
+     d <- subset(fev, subject %in% samp)
+     print(xyplot(log.fev1 ~ height, data = d, groups = subject,
+         type = "b", xlim = xl, ylim = yl))
+     Sys.sleep(0.5)
+ }
```

(a) Benutzen Sie die beiden festen Effekte Alter (`age`) und Körpergröße (`height`) in einem gemischten linearen Modell für die Lungenleistung. Verwenden Sie für die zufälligen Effekte subjektspezifische Intercepts und subjektspezifische Trends über die Zeit. Können die Einflüsse von `age` und `height` linear modelliert werden? Vergleichen Sie die Modellgüte verschiedener (Kombinationen von) Transformationen dieser beiden Größen.

(b) Benutzen Sie im Folgenden quadratische Trends für Alter und Körpergröße. Über prüfen Sie (grafisch), ob die Modell-Annahmen über die Residuen $\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i;\ i = 1, .., 300$ erfüllt sind.

   (i) Sind die Residuen für die einzelnen Subjekte in etwa symmetrisch und mit gleicher Varianz verteilt?

   (ii) Spiegelt das Modell die Form des Trends in Alter adäquat wieder oder gibt es systematische Unter-/Überschätzungen?

   (iii) Ist die Varianz der Residuen unabhängig vom Alter?

   (iv) Sind die Residuen normalverteilt?

   (v) Ist die Annahme unkorrelierter Residuen gerechtfertigt oder gibt es Hinweise auf das Vorliegen einer seriellen Korrelationsstruktur in `age`?

(c) Überprüfen Sie die Verteilungsannahme für die zufälligen Effekte.

(d) Erweitern Sie das Modell um eine serielle Korrelation und vergleichen Sie das erweiterte Modell mit dem ohne serielle Korrelationsstruktur. Probieren Sie verschiedene Typen von serieller Korrelation aus.

Hinweis: Falls der Schätzalgorithmus nicht konvergieren sollte, vereinfachen Sie die Kovarianzstruktur der zufälligen Effekte (`pdDiag`!) bevor Sie die serielle Korrelation mit ins Modell aufnehmen.

Aufgabe 2:

(a) Programmieren Sie eine R-Funktion, die die auf Folie 8.29 präsentierte Methode zur Identifikation von extremen Subjekten implementiert.

(b) Benutzen Sie die Funktion um auf Basis des End-Modells aus Aufgabe 1 Ausreißer-Subjekte zu identifizieren. Benutzen Sie $p < 0.01$ als Kriterium für Ausreißer.

(c) Überprüfen Sie den Einfluss dieser extremen Beobachtungen auf die Modellschätzung.