

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.

Lösung 1

```
(a) library(MASS)
library(lattice)
library(nlme)

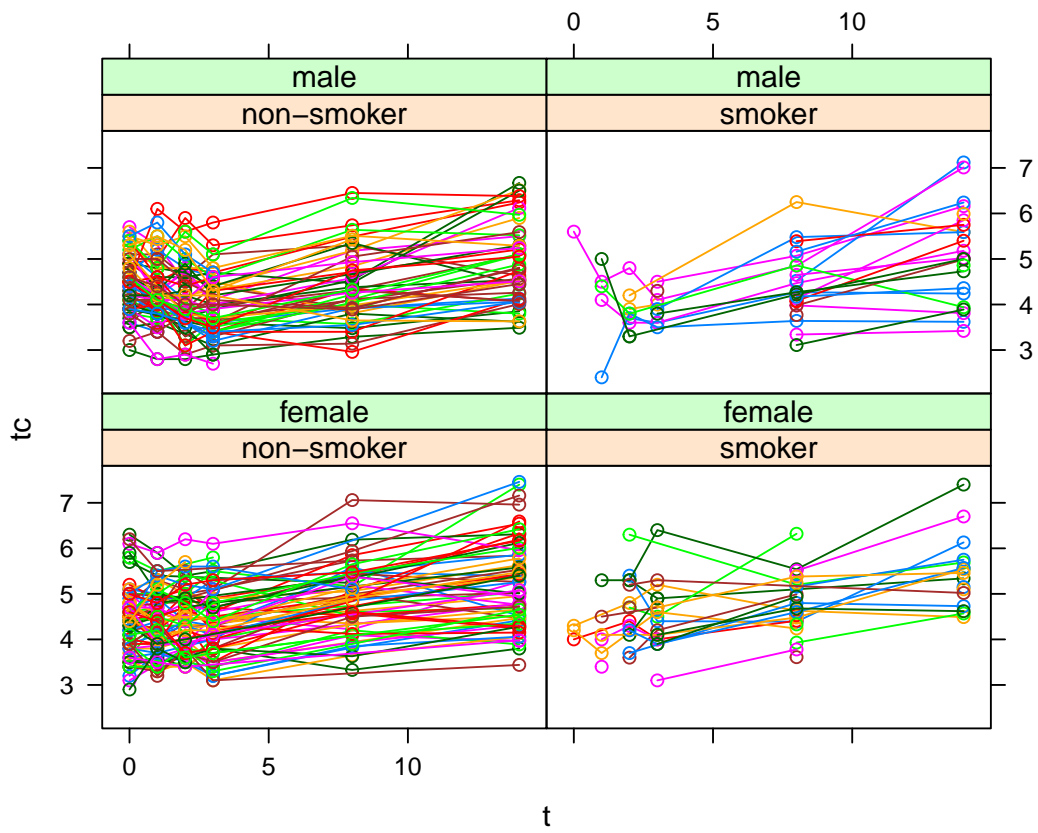
url <- paste0("https://www.elab.moodle.elearning.lmu.de/pluginfile.php/32320/",
              "course/section/12446/amsterdam.txt")
ex <- read.table(url, header=TRUE)

head(ex)

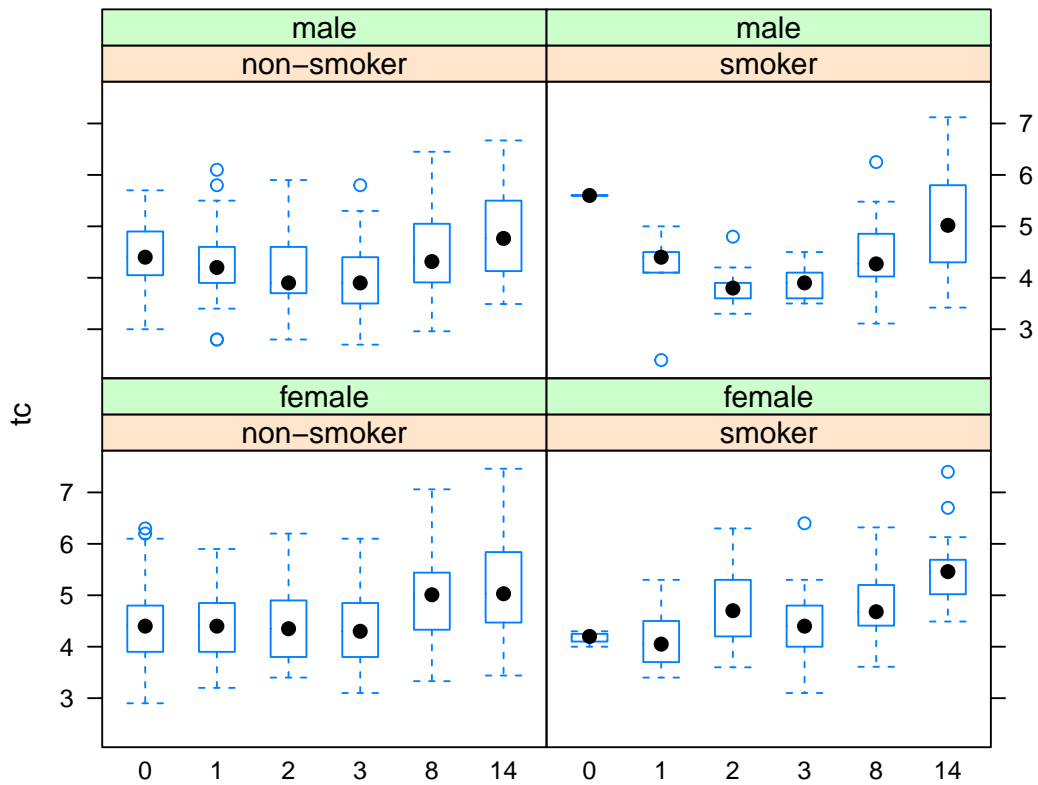
  id  tc  fitness bfatness smoker gender  t
1  1 4.20 2.151339    2.51  FALSE  MALE  0
2  1 3.90 2.151339    2.10  FALSE  MALE  1
3  1 3.90 2.151339    2.16  FALSE  MALE  2
4  1 3.60 2.151339    2.26  FALSE  MALE  3
5  1 3.92 2.151339    2.33  FALSE  MALE  8
6  1 4.13 2.151339    2.56  FALSE  MALE 14

ex$smoker <- factor(ex$smoker, labels=c("non-smoker", "smoker"))
ex$gender <- factor(ex$gender, labels=c("female", "male"))

## spaghetti plot of TC for the four combinations of smoker and gender
xyplot(tc ~ t | smoker*gender, groups=id, data=ex, type="b")
```



```
## boxplots of TC over time for the four combinations
bwplot(tc ~ as.factor(t) | smoker*gender, groups=id, data=ex)
```



```
## means and standard deviations of TC for both sexes
t <- unique(ex$t)
means <- with(ex,tapply(tc,list(t,gender),mean))
sds <- with(ex,tapply(tc,list(t,gender),sd))
res <- cbind(means[,1], sds[,1], means[,2], sds[,2])
colnames(res) <- c("mean(F)","sd(F)","mean(M)","sd(M)")
res
```

	mean(F)	sd(F)	mean(M)	sd(M)
0	4.408974	0.7263475	4.463768	0.6128456
1	4.379487	0.6665201	4.259420	0.6640603
2	4.434615	0.6982031	4.075362	0.6706901
3	4.370513	0.7222096	3.950725	0.6098754
8	4.880769	0.7230936	4.431739	0.7832098
14	5.252949	0.9068056	4.973043	0.9262405

(b)

$$TC_{ij} = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + \varepsilon_{ij}, \quad i = 1, \dots, 147; \quad j = 1, \dots, 6$$

$$\mathbf{b}_i = \begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{D}), \quad \mathbf{D} = \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{pmatrix}, \quad \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad \mathbf{b}_i, \varepsilon_{ij} \text{ unabhängig}$$

In R:

```
m.slope.lme <- lme(tc ~ 1 + t, random= ~ 1+t|id, data=ex)
# summary(m.slope.lme)

## parameters
fixef(m.slope.lme)

(Intercept)          t
4.22173897  0.05922892

m.slope.lme$sigma^2

[1] 0.1601557

getVarCov(m.slope.lme)

Random effects variance covariance matrix
              (Intercept)          t
(Intercept)  0.3354700  0.0016762
t            0.0016762  0.0014295
Standard Deviations: 0.5792 0.037809

# plot for fitted values
# plot(m.slope.lme, form=fitted(.) ~ t, type="b")
```

- (c) Kovarianzmatrix von Subjekt i : $V_i = Z_i D Z_i' + R_i$
Kovarianzmatrix gesamter Datensatz: $V = \text{blockdiag}(V_i) = Z G Z' + R$

```
## marginal covariance matrix for observations of one person
t <- unique(ex$t)
Zi <- cbind(1, t)
D <- getVarCov(m.slope.lme)
R <- diag(m.slope.lme$sigma^2, 6)
(Vi <- Zi %*% D %*% t(Zi) + R)

      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.4956261 0.3371465 0.3388227 0.3404989 0.3488797 0.3589367
[2,] 0.3371465 0.5004079 0.3433579 0.3464636 0.3619921 0.3806263
[3,] 0.3388227 0.3433579 0.5080488 0.3524284 0.3751045 0.4023159
[4,] 0.3404989 0.3464636 0.3524284 0.5185488 0.3882170 0.4240055
[5,] 0.3488797 0.3619921 0.3751045 0.3882170 0.6139348 0.5324536
```

```
[6,] 0.3589367 0.3806263 0.4023159 0.4240055 0.5324536 0.8227470
```

```
getVarCov(m.slope.lme, individuals="1", type="marginal")
```

```
id 1
```

```
Marginal variance covariance matrix
```

```
      1      2      3      4      5      6
1 0.49563 0.33715 0.33882 0.34050 0.34888 0.35894
2 0.33715 0.50041 0.34336 0.34646 0.36199 0.38063
3 0.33882 0.34336 0.50805 0.35243 0.37510 0.40232
4 0.34050 0.34646 0.35243 0.51855 0.38822 0.42401
5 0.34888 0.36199 0.37510 0.38822 0.61393 0.53245
6 0.35894 0.38063 0.40232 0.42401 0.53245 0.82275
```

```
Standard Deviations: 0.70401 0.7074 0.71278 0.7201 0.78354 0.90705
```

```
## covariance matrix for whole data set is block diagonal matrix
```

```
## with covariance matrices of each person on the diagonal
```

```
## example for n = 2 individuals
```

```
# require(Matrix)
```

```
# temp <- list(Vi = Vi)
```

```
# listVi <- rep(temp, 2)
```

```
# bdiag(listVi)
```

```
# correlation matrix for one subject
```

```
# cov2cor(Vi)
```

```
(d) cor.pos <- matrix(c(1,0.9,0.9,1),2,2)
```

```
cor.neg <- matrix(c(1,-0.9,-0.9,1),2,2)
```

```
library(mvtnorm)
```

```
# create plot with lines from intercepts and slopes
```

```
# random effects: (b_i1, b_i1)' ~ N( (0, 0)', corM),
```

```
# fixed effects: beta_0 = 5, beta_1 = 2
```

```
do.plot <- function(corM, main) {
```

```
  # simulate N( (5, 2)', corM)
```

```
  res <- mvrnorm(n=10, mu=c(5, 2), Sigma=corM)
```

```
  # plot effects
```

```
  plot(NA, xlim=c(0,5), ylim=c(3,10), type="n", xlab="t", ylab="", main=main)
```

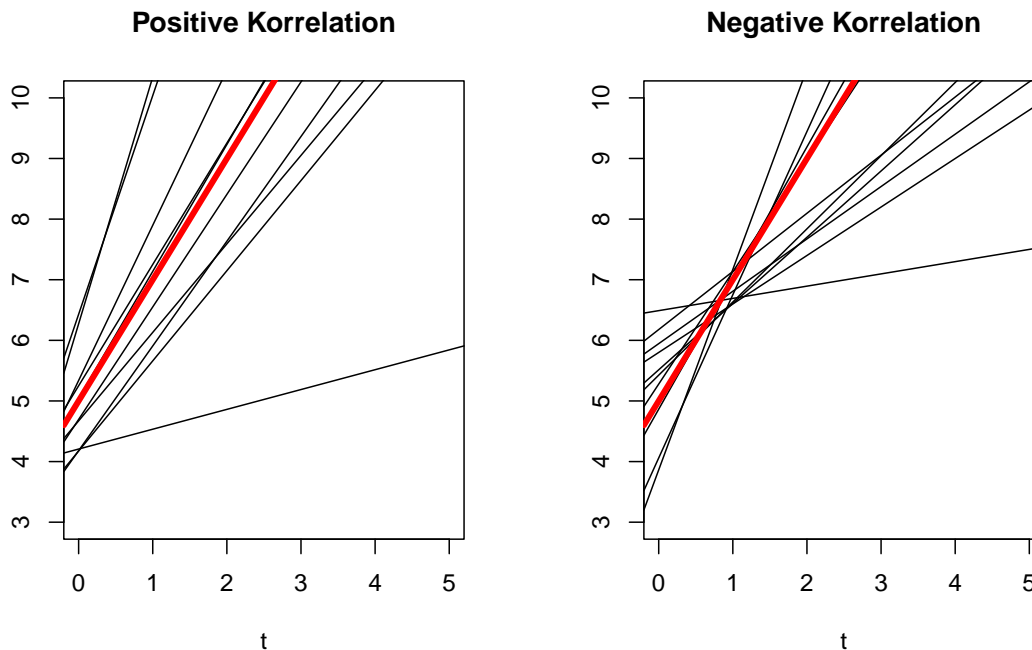
```
  for (i in 1:nrow(res)) abline(coef=res[i,])
```

```
  abline(coef=c(5,2), lty=1, lwd=4, col='red') # highlight fixed effects
```

```

}
x11(w=12, h=6)
par(mfcol=c(1,2))
set.seed(123)
do.plot(cor.pos, "Positive Korrelation")
do.plot(cor.neg, "Negative Korrelation")

```



(e)

$$\begin{aligned}
TC_{ij} &= \beta_0 + f_1(t)I(\text{sex}_i = \text{male}) + f_2(t)I(\text{sex}_i = \text{female}) + \beta_2 \text{bfatness}_{ij} + \beta_3 I(\text{smoker}_{ij} = \text{ja}) \\
&\quad + b_{i0} + b_{i1}t + \varepsilon_{ij} \\
&\quad i = 1, \dots, 147; \quad t \in \{0, 1, 2, 3, 8, 14\}
\end{aligned}$$

$$\mathbf{b}_i = \begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{D}), \quad \mathbf{D} = \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{pmatrix}, \quad \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad \mathbf{b}_i, \varepsilon_{ij} \text{ unabhängig}$$

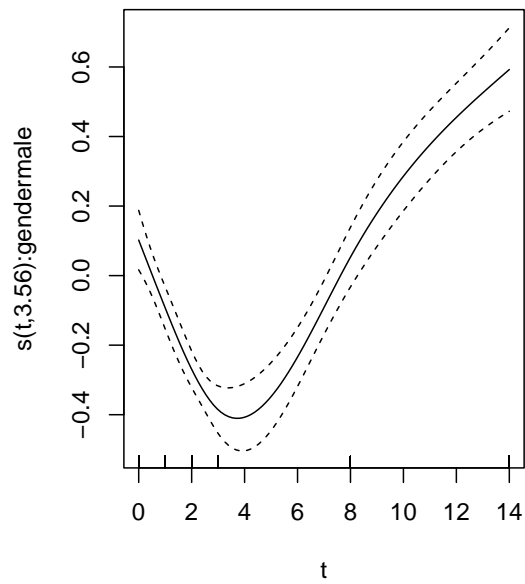
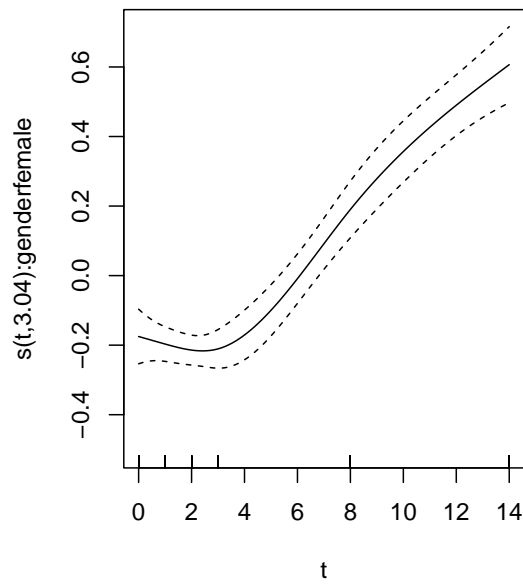
In R:

```

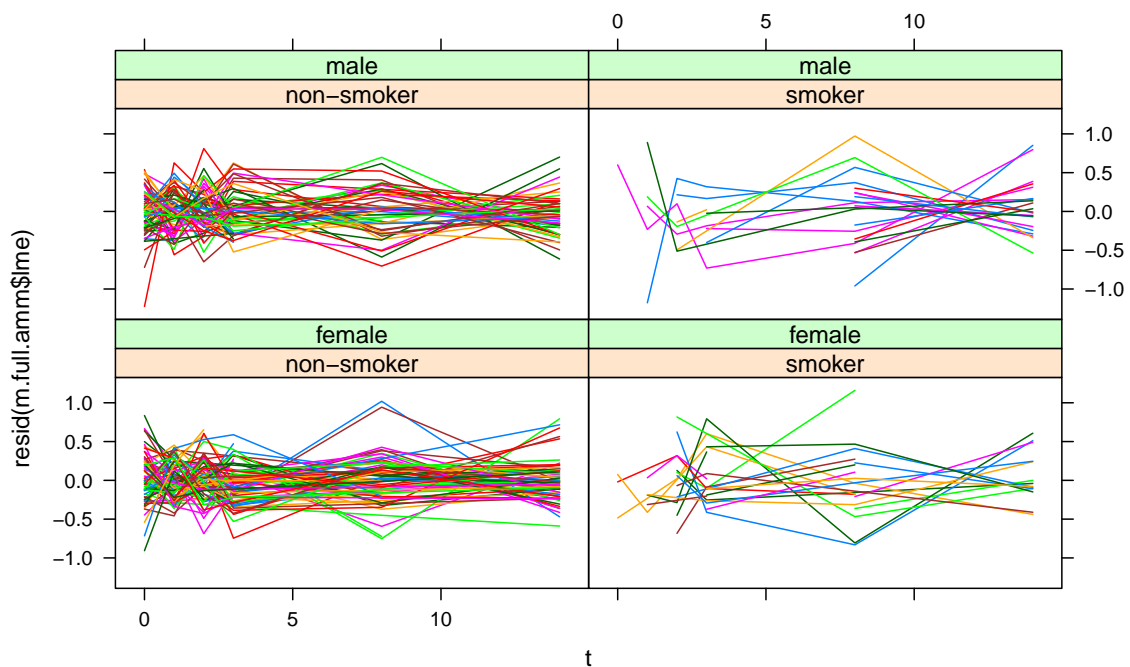
require(mgcv)
m.full.amm <- gamm(tc ~ 1 + s(t, by=gender, k=5) + bfatness + smoker,
                  random= list(id=~1+t), data=ex)
# summary(m.full.amm$gam)
# summary(m.full.amm$lme)

plot(m.full.amm$gam, pages=1)

```



```
# plot der Residuen
xyplot(resid(m.full.amm$lme) ~ t|smoker*gender, groups=id, data=ex, type="l")
```



```
# plot der Prognosen
xyplot(fitted(m.full.amm$lme) ~ t|smoker*gender, groups=id, data=ex, type="l")
```