Analyse Longitudinaler Daten SoSe 2016 **Old exam questions**

Prof. Dr. Sonja Greven und Dipl. Stat. Jona Cederbaum 23.06.2016

Note: This is an extract from old exams which were 120 minutes long and in which 100 points could be scored. The exam in this term (Semester) will take only 90 minutes.

Exercise 1:

In the past years, there were many so-called "summers of the century" in which the hottest temperatures were measured since recording began. For some people, these high temperatures lead to health problems. A study will investigate if a continuously consumption of salt and minerals can lead to improvement.

Weekly measurements of the temperature (`temp`) in the summer months were taken and the blood pressure (`pressure`) of each study participant was recorded as an indicator for the health status. The participants were assigned to a treatment and to a placebo group (`group`, 0:placebo, 1:treatment). In addition, the age at the beginning of the study (`age`, average: 44 years) of each participant was collected.

Your statistics team fits the following model to the data:

```
m1 <- lme(pressure ~ age + time * group + temp, random = ~ 1 + time | id, method = 'REML').
```

a) (13P) Formulate the underlying model **for this specific application** and name/describe **all** model components (response, effects, indices, etc.). Also write down **all** assumptions of the specified model.

b) (2P) The study director claims that the assignment to the groups took place randomly. You have doubts. Formulate a suitable null hypothesis to check this and name an appropriate test.

c) (5P) It seems plausible to you that the blood pressure of the participants strongly varies at the beginning of the study. In order to check if the subject-specific slopes are also justified, you perform the following test:

```
m2 <- update(m1, random = ~ 1 | id)
anova(m1,m2).
```

The p-value given out by `anova()` is greater than the predefined significance level. What conclusions do you draw from this for the inclusion of the subject-specific slopes? What would be your answer if the p-value was smaller than the significance level? Justify both of your answers.

d) (2P) The model diagnosis shows that a linear time trend is not appropriate here. How could you include both a smooth time trend as well as random effects in one model? Which `R`-function can you use for this?

Exercise 2: (5P)

Show (using formulas) that the fixed effects in the generalized linear mixed model generally cannot be interpreted on population level.

Exercise 3:

In the following, you can see the beginning of two different statements (A-B). Complete each sentence with one of the three given options. For each sentence, exactly one option is correct. Mark the correct option on your exam copy (no justification necessary). For each correct answer, you obtain 2 points, for each incorrect answer one minus point. At most 4 and at least 0 points can be obtained in this problem set.

A. The conditional AIC (cAIC) differs from the marginal AIC (mAIC) through the fact that

   (1) the cAIC is based on maximum likelihood (ML), the mAIC, however, is based on restricted maximum likelihood (REML).

   (2) the mAIC is also appropriate for REML, the cAIC is not.

   (3) the mAIC is based on predictions for replications with different random effects, the cAIC, however, is based on predictions for replications with the same random effects.

B. Consider the following output of a linear mixed model:

```
Linear mixed-effects model fit by REML
Data: penguin
AIC      BIC    logLik
945.3361 976.8841 -463.6681

Random effects:
  Formula: ~1 + time | id
Structure: Diagonal
(Intercept)        time Residual
StdDev:     1.87764 0.000123266 1.201899

Fixed effects: y ~ time * group
Value Std.Error  DF   t-value p-value
(Intercept)       69.04064 0.5472666 199 126.15541  0.0000
time               7.40359 0.2482171 199  29.82709  0.0000
grouphigh         -1.15059 0.7839434  47  -1.46769  0.1488
groupcontrol      -0.12192 0.8211320  47  -0.14847  0.8826
time:grouphigh    -0.36478 0.3531854 199  -1.03282  0.3029
time:groupcontrol -0.17640 0.4012671 199  -0.43961  0.6607
Correlation:
  (Intr) time   grphgh grpcnt tm:grph
time              -0.530
grouphigh         -0.698  0.370
groupcontrol      -0.666  0.353  0.465
time:grouphigh     0.372 -0.703 -0.528 -0.248
time:groupcontrol  0.328 -0.619 -0.229 -0.536  0.435
```

```
Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-2.27886861 -0.65398955 -0.01280745  0.55844581  2.84425395

Number of Observations: 252
Number of Groups: 50
```

You can conclude from the output that

(1) the random effects were assumed to be uncorrelated.

(2) the observations on one subject were assumed to be uncorrelated.

(3) the fixed effects were assumed to be uncorrelated.

Exercise 4: (3P)

In an epilepsy study, the absence/presence of epileptic seizures (`seizure`, 0: absence, 1: presence) of 100 patients was measured monthly after initial hospitalization into a special clinic. The sex (`sex`, 0: women, 1: men) of the patients was additionally collected. In the following, the estimated model is shown.

```
require(lme4)
glmer(seizure ~ month + sex + (1 | subject), family = binomial(), data = epil, nAGQ = 1)
```

You obtain among other things the following model output.

```
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.46713    0.40034   1.167   0.2433
month       -0.54208    0.04453 -12.174   <2e-16 ***
sex          1.21287    0.53115   2.284   0.0224 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpret the estimate of the fixed effect of `month`.

Exercise 5:

It is generally known that a variety of bacteria are already resistant against some antibiotics. Therefore, new types of antibiotics are developed. For this, a team of experts compares two new preparations (Med1, Med2) with a commonly used antibiotic (Med3).

The 300 study participants, for whom the same bacterium shall be fought against, are randomized into the three groups. Their state of health is measured in form of a metric score. The first measurement takes place at the beginning of the study. All participants come together to the first measurement. More collective measurements follow after 2,4,6 and 7 days, where all participants are present.

a) (2P) Are the data in their present form equidistant? Justify your answer.

b) (3P) The following plot shows the course of the score over time by groups. Based on this plot, explain what seems do have gone wrong in the study. How do you account for this in your subsequent modeling?