5 Messwiederholungen - Random Effects Modelle

Aufgabe 26

Analyse des Datensatz 'ohio'

```
# Daten
library(geepack)
data(ohio)
ohio[1:30,]

##    resp id age smoke
## 1     0  0  -2     0
## 2     0  0  -1     0
## 3     0  0   0     0
## 4     0  0   1     0
## 5     0  1  -2     0
## 6     0  1  -1     0
## 7     0  1   0     0
## 8     0  1   1     0
## 9     0  2  -2     0
## 10    0  2  -1     0
## 11    0  2   0     0
## 12    0  2   1     0
## 13    0  3  -2     0
## 14    0  3  -1     0
## 15    0  3   0     0
## 16    0  3   1     0
## 17    0  4  -2     0
## 18    0  4  -1     0
## 19    0  4   0     0
## 20    0  4   1     0
## 21    0  5  -2     0
## 22    0  5  -1     0
## 23    0  5   0     0
## 24    0  5   1     0
## 25    0  6  -2     0
## 26    0  6  -1     0
## 27    0  6   0     0
## 28    0  6   1     0
## 29    0  7  -2     0
## 30    0  7  -1     0
```

(a) GLM mit Probit-Link

```
glm1 <- glm(resp ~ age*smoke, data=ohio, family=binomial(link=probit))
summary(glm1)

##
## Call:
## glm(formula = resp ~ age * smoke, family = binomial(link = probit),
##     data = ohio)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6507  -0.6014  -0.5640  -0.4932   2.0818
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.12594    0.04712 -23.896   <2e-16 ***
## age         -0.07681    0.03749  -2.049   0.0405 *
## smoke        0.17088    0.07612   2.245   0.0248 *
## age:smoke    0.03673    0.06110   0.601   0.5477
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1829.1  on 2147  degrees of freedom
## Residual deviance: 1819.4  on 2144  degrees of freedom
## AIC: 1827.4
##
## Number of Fisher Scoring iterations: 4
```

(b) Modell mit random intercept

```
library(lme4)
glmer1 <- glmer(resp ~ age*smoke + (1| id), family=binomial(link=probit),
data=ohio, nAGQ=20)
summary(glmer1)

## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 20) [glmerMod]
##  Family: binomial  ( probit )
## Formula: resp ~ age * smoke + (1 | id)
##    Data: ohio
##
##      AIC      BIC   logLik deviance df.resid
##   1605.3   1633.7   -797.7   1595.3     2143
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2955 -0.1838 -0.1590 -0.1170  2.4915
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  id     (Intercept) 1.491    1.221
## Number of obs: 2148, groups:  id, 537
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.76679    0.12088 -14.616   <2e-16 ***
## age         -0.12271    0.04815  -2.549   0.0108 *
## smoke        0.25418    0.15872   1.601   0.1093
## age:smoke    0.06075    0.07788   0.780   0.4354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) age    smoke
## age        0.299
## smoke     -0.522 -0.195
## age:smoke -0.170 -0.616  0.288
```

Beachte: betas sind unterschiedlich zu interpretieren

- GLMM: subjektspezifisch
- GLM: populationsspezifisch

Hintergrund:

- GLMM: Probitmodell für bedingte Verteilung $y_{it}|x_{it}, b_i$
- GLM: Probitmodell für marginale Verteilung $y_{it}|x_{it}$

Man erkennt deutlich den Shrinkage-Effekt bei geschätzten betas im GLM verglichen zum Random-Intercept-Modell. Im Falle des Probit-Normal-Modells mit Random Intercept lässt sich Shrinkage-Effekt sogar formalisieren

```
sigmab2 <- VarCorr(glmer1)$id[1,1]
betavector <- fixef(glmer1)
betavector/sqrt(1 + sigmab2)

## (Intercept)         age       smoke   age:smoke
## -1.11937221 -0.07774727  0.16103786  0.03848912

coef(glm1)

## (Intercept)         age       smoke   age:smoke
## -1.12594080 -0.07680844  0.17088443  0.03673144
```

Stimmt hier nur einigermaßen; Grund: Approximation des Integrals bei glmer: Gauß-Hermite-Quadratur (adaptive Wahl der Stützstellen)

(c) GLM mit Logit-Link

```
glm2 <- glm(resp ~ age*smoke, data=ohio, family=binomial(link=logit))
summary(glm2)

##
## Call:
## glm(formula = resp ~ age * smoke, family = binomial(link = logit),
##     data = ohio)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6503  -0.6014  -0.5636  -0.4940   2.0804
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.90084    0.08874 -21.420   <2e-16 ***
## age         -0.14125    0.06951  -2.032   0.0422 *
## smoke        0.31395    0.13944   2.252   0.0244 *
## age:smoke    0.07084    0.11072   0.640   0.5223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1829.1  on 2147  degrees of freedom
## Residual deviance: 1819.5  on 2144  degrees of freedom
## AIC: 1827.5
##
## Number of Fisher Scoring iterations: 4
```

(d) Modelle mit **glmmML**

```
library(glmmML)
?glmmML
# nur Random Intercept Modelle möglich
```

ML mit Gauß-Hermite-Quadratur

```
ghq20 <- glmmML(resp ~ age * smoke, method = "ghq", cluster = id, data = ohio, n.points = 20)
summary(ghq20)
```

```
##
## Call:  glmmML(formula = resp ~ age * smoke, data = ohio, cluster = id,      method = "ghq", n.po
##
##
##                coef se(coef)        z Pr(>|z|)
## (Intercept) -3.1282  0.22284 -14.0381   0.0000
## age         -0.2164  0.08656  -2.4998   0.0124
## smoke        0.4620  0.28555   1.6179   0.1060
## age:smoke    0.1053  0.13850   0.7606   0.4470
##
## Scale parameter in mixing distribution:  2.166 gaussian
## Std. Error:                              0.1852
##
##         LR p-value for H_0: sigma = 0:  4.145e-51
##
## Residual deviance: 1595 on 2143 degrees of freedom  AIC: 1605
```

```
ghq15 <- glmmML(resp ~ age * smoke, method = "ghq", cluster = id, data = ohio, n.points = 15)
summary(ghq15)
```

```
##
## Call:  glmmML(formula = resp ~ age * smoke, data = ohio, cluster = id,      method = "ghq", n.po
##
##
##                coef se(coef)        z Pr(>|z|)
## (Intercept) -3.1266  0.22225 -14.0682   0.0000
## age         -0.2167  0.08656  -2.5037   0.0123
## smoke        0.4618  0.28533   1.6185   0.1060
## age:smoke    0.1053  0.13849   0.7602   0.4470
##
## Scale parameter in mixing distribution:  2.164 gaussian
## Std. Error:                              0.1843
##
##         LR p-value for H_0: sigma = 0:  4.218e-51
##
## Residual deviance: 1595 on 2143 degrees of freedom  AIC: 1605
```

```
ghq10 <- glmmML(resp ~ age * smoke, method = "ghq", cluster = id, data = ohio, n.points = 10)
summary(ghq10)
```

```
##
## Call:  glmmML(formula = resp ~ age * smoke, data = ohio, cluster = id,      method = "ghq", n.po
##
##
##                coef se(coef)       z Pr(>|z|)
## (Intercept) -3.1163  0.21940 -14.204   0.0000
## age         -0.2160  0.08648  -2.498   0.0125
## smoke        0.4613  0.28415   1.623   0.1040
```

```
## age:smoke    0.1052  0.13836   0.760   0.4470
##
## Scale parameter in mixing distribution:  2.15 gaussian
## Std. Error:                              0.1792
##
##         LR p-value for H_0: sigma = 0:  4.727e-51
##
## Residual deviance: 1595 on 2143 degrees of freedom  AIC: 1605
```

```
ghq5 <- glmmML(resp ~ age * smoke, method = "ghq", cluster = id, data = ohio, n.points = 5)
summary(ghq5)
```

```
##
## Call:  glmmML(formula = resp ~ age * smoke, data = ohio, cluster = id,      method = "ghq", n.po
##
##
##                coef se(coef)        z Pr(>|z|)
## (Intercept) -3.0500   0.2073 -14.7159   0.0000
## age         -0.2134   0.0859  -2.4837   0.0130
## smoke        0.4566   0.2756   1.6569   0.0975
## age:smoke    0.1039   0.1375   0.7555   0.4500
##
## Scale parameter in mixing distribution:  2.05 gaussian
## Std. Error:                              0.1586
##
##         LR p-value for H_0: sigma = 0:  2.33e-50
##
## Residual deviance: 1598 on 2143 degrees of freedom  AIC: 1608
```

ML mit Laplace-Approximation (= Gauß-Hermite-Quadratur mit n.points=1)

```
laplace1 <- glmmML(resp ~ age * smoke, method = "Laplace", cluster = id, data = ohio)
summary(laplace1)
```

```
##
## Call:  glmmML(formula = resp ~ age * smoke, data = ohio, cluster = id,      method = "Laplace")
##
##
##                coef se(coef)        z Pr(>|z|)
## (Intercept) -3.4017   0.3024 -11.2474   0.0000
## age         -0.2170   0.0869  -2.4977   0.0125
## smoke        0.4782   0.2995   1.5970   0.1100
## age:smoke    0.1046   0.1391   0.7521   0.4520
##
## Scale parameter in mixing distribution:  2.346 gaussian
## Std. Error:                              0.2707
##
##         LR p-value for H_0: sigma = 0:  2.741e-52
##
## Residual deviance: 1589 on 2143 degrees of freedom  AIC: 1599
```

```
laplace2 <- glmmML(resp ~ age * smoke, method = "ghq", cluster = id, data = ohio,
    n.points = 1)
summary(laplace2)
```

```
##
## Call:  glmmML(formula = resp ~ age * smoke, data = ohio, cluster = id,      method = "ghq", n.po
##
##
```

```
##                 coef se(coef)        z Pr(>|z|)
## (Intercept) -3.4017   0.3024 -11.2474   0.0000
## age         -0.2170   0.0869  -2.4977   0.0125
## smoke        0.4782   0.2995   1.5970   0.1100
## age:smoke    0.1046   0.1391   0.7521   0.4520
##
## Scale parameter in mixing distribution:  2.346 gaussian
## Std. Error:                              0.2707
##
##         LR p-value for H_0: sigma = 0:  2.741e-52
##
## Residual deviance: 1589 on 2143 degrees of freedom  AIC: 1599
```

→ Ergebnisse werden von der Anzahl der Stützstellen beeinflusst

(e) Modelle mit `lme4`

```
library(lme4)
```

ML mit Gauß-Hermite-Quadratur (mit adaptiver Wahl der Stütztstellen)

```
glmer2 <- glmer(resp ~ age * smoke + (age | id), family="binomial", data=ohio, nAGQ=20)

## Error in updateGlmerDevfun(devfun, glmod$reTrms, nAGQ = nAGQ): nAGQ > 1 is only
available for models with a single, scalar random-effects term
```

ML mit Laplace-Approximation (= Gauß-Hermite-Quadratur mit nAGQ=1)

```
glmer3 <- glmer(resp ~ age * smoke + (age | id), family="binomial", data=ohio, nAGQ=1)
summary(glmer3)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: resp ~ age * smoke + (age | id)
##    Data: ohio
##
##      AIC      BIC   logLik deviance df.resid
##   1601.3   1641.1   -793.7   1587.3     2141
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.24838 -0.18153 -0.14760 -0.09758  2.18559
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  id     (Intercept) 7.0333   2.6520
##         age         0.2536   0.5036   0.45
## Number of obs: 2148, groups:  id, 537
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.7449     0.4285  -8.739   <2e-16 ***
## age          -0.3871     0.2778  -1.393    0.164
## smoke         0.5102     0.3275   1.558    0.119
## age:smoke     0.1259     0.1566   0.804    0.421
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) age    smoke
## age        0.299
## smoke     -0.357 -0.134
## age:smoke -0.188 -0.346  0.397
```

(f) Modelle mit `MASS`

```
library(MASS)
```

Mit Penalized Quasi-Likelihood (als Laplace-Approximation motivierbar)

```
pql1 <- glmmPQL(resp ~ age * smoke , random = ~age | id, data=ohio,
 family=binomial(link=logit))
summary(pql1)

## Linear mixed-effects model fit by maximum likelihood
##  Data: ohio
##   AIC BIC logLik
##    NA  NA     NA
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 2.5176745 (Intr)
## age         1.1928183 0.291
## Residual    0.5122051
##
## Variance function:
##  Structure: fixed weights
##  Formula: ~invwt
## Fixed effects: resp ~ age * smoke
##                  Value Std.Error   DF    t-value p-value
## (Intercept) -3.312941 0.1746951 1609 -18.964130  0.0000
## age         -0.296199 0.1015284 1609  -2.917400  0.0036
## smoke        0.518257 0.2853085  535   1.816478  0.0699
## age:smoke    0.157451 0.1648446 1609   0.955147  0.3396
##  Correlation:
##           (Intr) age    smoke
## age        0.406
## smoke     -0.612 -0.249
## age:smoke -0.250 -0.616  0.376
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.8104040 -0.2677280 -0.2369320 -0.1855579  3.3096332
##
## Number of Observations: 2148
## Number of Groups: 537
```

(g) siehe Übungsmitschrift