

6 Marginale Modelle

Aufgabe 1

Die Daten

```
library(MASS)
data(epil)
#help(epil)
```

a) GLM mit Poissonverteiltem y

```
glm1 <- glm(y ~ lbase + trt + lage + V4 + lbase:trt, family=poisson, data=epil)
summary(glm1)

##
## Call:
## glm(formula = y ~ lbase + trt + lage + V4 + lbase:trt, family = poisson,
##      data = epil)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.0915  -1.4126  -0.2739   0.7580  10.7711
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.89791    0.04260  44.552 < 2e-16 ***
## lbase           0.94862    0.04360  21.759 < 2e-16 ***
## trtprogabide   -0.34588    0.06100  -5.670 1.42e-08 ***
## lage           0.88760    0.11650   7.619 2.56e-14 ***
## V4            -0.15977    0.05458  -2.927 0.00342 **
## lbase:trtprogabide 0.56154    0.06352   8.841 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2517.83  on 235  degrees of freedom
## Residual deviance:  869.07  on 230  degrees of freedom
## AIC: 1647
##
## Number of Fisher Scoring iterations: 5
```

Nach den geschätzten Koeffizienten zu urteilen reduziert progabide die erwartete Anzahl an Anfällen, allerdings nur für weniger schwerwiegende Fälle (niedriges oder mäßiges lbase). Für hohes lbase ist der Zusammenhang dem Anschein nach umgekehrt (positiver Koeffizient beim Wechselwirkungsterm!). Bei älteren Patienten ist die Anzahl der Anfälle offenbar höher, abenso (wie zu erwarten) bei einer großen Anzahl in der Baseline-Periode. In der vierten Woche scheinen die Anfälle weniger zu werden.

Problematisch an dem GLM ist jedoch: Die einzelnen Beobachtungen sind nicht unabhängig, da jeder Patient viermal eingeht.

```
table(epil$subject,epil$period)[1:12,]
```

```
##
##      1 2 3 4
##  1  1 1 1 1
##  2  1 1 1 1
##  3  1 1 1 1
##  4  1 1 1 1
##  5  1 1 1 1
##  6  1 1 1 1
##  7  1 1 1 1
##  8  1 1 1 1
##  9  1 1 1 1
## 10  1 1 1 1
## 11  1 1 1 1
## 12  1 1 1 1
```

Die Messungen an einer Person sind mit Sicherheit abhängig/positiv korreliert. Daher sind insbesondere die geschätzten Standardfehler und somit die angegebenen Teststatistiken nicht zuverlässig.

b) Quasi-Poisson-Modell

```
glm2 <- glm(y ~ lbase + trt + lage + V4 + lbase:trt, family=quasipoisson, data=epil)
summary(glm2)
```

```
##
## Call:
## glm(formula = y ~ lbase + trt + lage + V4 + lbase:trt, family = quasipoisson,
##      data = epil)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.0915  -1.4126  -0.2739   0.7580  10.7711
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.89791    0.08950  21.206 < 2e-16 ***
## lbase           0.94862    0.09159  10.357 < 2e-16 ***
## trtprogabide   -0.34588    0.12815  -2.699 0.007471 **
## lage           0.88760    0.24475   3.627 0.000354 ***
## V4            -0.15977    0.11468  -1.393 0.164897
## lbase:trtprogabide 0.56154    0.13345   4.208 3.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.413874)
##
##      Null deviance: 2517.83  on 235  degrees of freedom
## Residual deviance:  869.07  on 230  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Da nur ein Dispersionsparameter eingeführt wurde, bleiben die Koeffizientenschätzer gleich. Der geschätzte Dispersionsparameter ist allerdings (mit 4.4) deutlich größer als 1. Es liegt also Overdispersion vor (vermutlich auf Grund der positiven Korrelation zwischen den Beobachtungen). Daher sind die geschätzten Standardfehler nun deutlich größer, was Auswirkungen auf die angegebene Teststatistiken hat (z.B. ist der Effekt von V4 nun nicht mehr signifikant).

c) Siehe Übungsmitschrift

d) GEEs

```
library(gee)
# help(gee)
```

Mit working correlation „independence“

```
gee1 <- gee(y ~ lbase + trt + lage + V4 + lbase:trt, id=subject, family=poisson,
corstr="independence", data=epil)
```

```
##          (Intercept)          lbase          trtprogabide          lage
##          1.8979148          0.9486222          -0.3458752          0.8875953
##                V4 lbase:trtprogabide
##          -0.1597696          0.5615356
```

```
summary(gee1)
```

```
##
## GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
## gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
## Link:                      Logarithm
## Variance to Mean Relation: Poisson
## Correlation Structure:     Independent
##
## Call:
## gee(formula = y ~ lbase + trt + lage + V4 + lbase:trt, id = subject,
##      data = epil, family = poisson, corstr = "independence")
##
## Summary of Residuals:
##          Min          1Q          Median          3Q          Max
## -14.3007316 -2.8918542 -0.4997212   1.7352535  59.8901336
##
##
## Coefficients:
##              Estimate Naive S.E.   Naive z Robust S.E.   Robust z
## (Intercept)   1.8979148 0.08949825  21.206165   0.11016938  17.227244
## lbase         0.9486222 0.09159327  10.356899   0.09648692   9.831614
## trtprogabide -0.3458752 0.12815011  -2.698985   0.17820422  -1.940892
## lage         0.8875953 0.24475026   3.626535   0.27273989   3.254366
## V4          -0.1597696 0.11467609  -1.393225   0.06514075  -2.452683
## lbase:trtprogabide 0.5615356 0.13344644   4.207948   0.17389100   3.229239
##
## Estimated Scale Parameter: 4.413871
## Number of Iterations: 1
##
## Working Correlation
##          [,1] [,2] [,3] [,4]
## [1,]      1   0   0   0
## [2,]      0   1   0   0
## [3,]      0   0   1   0
## [4,]      0   0   0   1
```

Mit equi-correlation, d.h. „exchangeable“

```
gee2 <- gee(y ~ lbase + trt + lage + V4 + lbase:trt, id=subject, family=poisson,
corstr="exchangeable", data=epil)
```

```
##          (Intercept)                lbase      trtprogabide                lage
##          1.8979148                0.9486222      -0.3458752                0.8875953
##          V4 lbase:trtprogabide
##          -0.1597696                0.5615356
```

```
summary(gee2)
```

```
##
## GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
## gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
## Link:                Logarithm
## Variance to Mean Relation: Poisson
## Correlation Structure: Exchangeable
##
## Call:
## gee(formula = y ~ lbase + trt + lage + V4 + lbase:trt, id = subject,
##      data = epil, family = poisson, corstr = "exchangeable")
##
## Summary of Residuals:
##      Min      1Q   Median      3Q      Max
## -14.306798 -2.898214 -0.489576  1.729516  59.915192
##
##
## Coefficients:
##              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
## (Intercept)   1.8949186 0.12458116 15.210315  0.11222853 16.884465
## lbase         0.9494588 0.13157877  7.215897  0.09865387  9.624142
## trtprogabide -0.3415598 0.18389541 -1.857359  0.18022069 -1.895231
## lage         0.8965103 0.35121890  2.552568  0.27506465  3.259271
## V4           -0.1597696 0.09229208 -1.731130  0.06514075 -2.452683
## lbase:trtprogabide 0.5625270 0.19151996  2.937172  0.17490853  3.216121
##
## Estimated Scale Parameter: 4.416317
## Number of Iterations: 2
##
## Working Correlation
##      [,1] [,2] [,3] [,4]
## [1,] 1.0000000 0.3542713 0.3542713 0.3542713
## [2,] 0.3542713 1.0000000 0.3542713 0.3542713
## [3,] 0.3542713 0.3542713 1.0000000 0.3542713
## [4,] 0.3542713 0.3542713 0.3542713 1.0000000
```

Zusätzlich autoregressiv (erster Ordnung), d.h. mit „AR-M“ (Mv=1)

```
gee3 <- gee(y ~ lbase + trt + lage + V4 + lbase:trt, id=subject, family=poisson,
corstr="AR-M", Mv=1, data=epil)
```

```
##          (Intercept)                lbase      trtprogabide                lage
##          1.8979148                0.9486222      -0.3458752                0.8875953
##          V4 lbase:trtprogabide
##          -0.1597696                0.5615356
```

```
summary(gee3)
```

```
##
## GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
## gee S-function, version 4.13 modified 98/01/27 (1998)
```

```

##
## Model:
## Link:                               Logarithm
## Variance to Mean Relation: Poisson
## Correlation Structure:      AR-M , M = 1
##
## Call:
## gee(formula = y ~ lbase + trt + lage + V4 + lbase:trt, id = subject,
##      data = epil, family = poisson, corstr = "AR-M", Mv = 1)
##
## Summary of Residuals:
##      Min      1Q      Median      3Q      Max
## -14.1246515 -2.9195050 -0.5511584  1.6995037  59.7165810
##
##
## Coefficients:
##              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
## (Intercept)      1.9050058 0.12217641 15.592255  0.10999427 17.319137
## lbase            0.9437141 0.12717789  7.420426  0.09271937 10.178176
## trtprogabide     -0.3871719 0.17994105 -2.151660  0.17169542 -2.254993
## lage            0.9835435 0.34231623  2.873201  0.27220893  3.613193
## V4              -0.1524001 0.09621486 -1.583956  0.08871766 -1.717811
## lbase:trtprogabide 0.6188674 0.18612510  3.325008  0.16924753  3.656582
##
## Estimated Scale Parameter:  4.465076
## Number of Iterations:  3
##
## Working Correlation
##      [,1] [,2] [,3] [,4]
## [1,] 1.0000000 0.4669392 0.2180322 0.1018078
## [2,] 0.4669392 1.0000000 0.4669392 0.2180322
## [3,] 0.2180322 0.4669392 1.0000000 0.4669392
## [4,] 0.1018078 0.2180322 0.4669392 1.0000000

```

Bei „independence“ ergeben sich für die Koeffizienten die gleichen Schätzwerte wie im GLM, bei „exchangable“ zeigen sich hier (Spezialfall!) nur numerische Unterschiede. Auch bei „AR-1“ sind sie qualitativ ähnlich.

Quantifizierung des Einflusses z.B. über

```
exp(gee3$coef["V4"])
```

```
##      V4
## 0.8586446
```

```
# d.h. in der vierten Woche nimmt die Anzahl der Anfälle marginal um den
# Faktor 0.86 ab.
```

Die Spalte „Naive z“ gibt den Standardfehler aus unter der Annahme, dass die verwendete (working) covariance der wahren entspricht (mit Dispersion). „Robust z“ verwendet die Schätzung über die Sandwich-Matrix.