# Knee Injuries - Marginal Models

February 5, 2020

First the dataset knee is loaded:

```
> library(catdata)
> data(knee)
> attach(knee)
```

To obtain a simple binary model the response variables are dichotomized. The groups are constructed by pain level up to level 2 und pain level higher than level 2.

```
> R2D <- rep(0, length(R2))
> R3D <- rep(0, length(R3))
> R4D <- rep(0, length(R3))
> R2D[R2>2] <- 1
> R3D[R3>2] <- 1
> R4D[R4>2] <- 1
```

Now the covariates have to be transformed so that they can be used for the functions "gee" from the "gee"–library and "geeglm" from the "geepack"–library, which will be employed for fitting the models.

```
> N <- rep(knee$N, each=3)
> Th <- rep(knee$Th, each=3)
> Age <- rep(knee$Age, each=3)
> Sex <- rep(knee$Sex, each=3)
```

Now the response vector is built and the quadratic age–effect "Age2" is computed.

```
> Response <- c(rbind(R2D,R3D,R4D))
> Age2 <- Age^2
```

The covariates therapy and sex are treated as factors:

```
> Th <- as.factor(Th)
> Sex <- as.factor(Sex)
```

First the GEEs are fitted with the funtion "gee" from library "gee".

```
> library(gee)
```

The first model is a GEE with independent correlation structure:

```
> gee1a <- gee(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit))

> summary(gee1a)

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

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent

Call:
gee(formula = Response ~ Th + Sex + Age + Age2, id = N, family = binomial(link = logit))

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8093879 -0.4771034  0.2393885  0.3881349  0.7660757


Coefficients:
               Estimate  Naive S.E.   Naive z Robust S.E.   Robust z
(Intercept) -4.749831307 1.279009755 -3.713679 1.821208734 -2.6080653
Th2         -0.673979165 0.223825316 -3.011184 0.334205312 -2.0166620
Sex1         0.265689239 0.241933592  1.098191 0.366733338  0.7244753
Age          0.381150842 0.087570504  4.352503 0.125833185  3.0290169
Age2        -0.006124345 0.001381627 -4.432705 0.001984839 -3.0855622

Estimated Scale Parameter:  1.019152
Number of Iterations:  1

Working Correlation
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

The second model is a GEE with exchangeable correlation structure:

```
> gee2a <- gee(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit), corstr="exchangeable")

> summary(gee2a)

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

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
```

```
 Correlation Structure:     Exchangeable

Call:
gee(formula = Response ~ Th + Sex + Age + Age2, id = N, family = binomial(link = logit),
    corstr = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8093879 -0.4771034  0.2393885  0.3881349  0.7660757


Coefficients:
               Estimate  Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -4.749831307 1.903793545 -2.4949298 1.821208734 -2.6080653
Th2         -0.673979165 0.333161800 -2.0229785 0.334205312 -2.0166620
Sex1         0.265689239 0.360115792  0.7377884 0.366733338  0.7244753
Age          0.381150842 0.130347841  2.9241055 0.125833185  3.0290169
Age2        -0.006124345 0.002056538 -2.9779873 0.001984839 -3.0855622

Estimated Scale Parameter:  1.019152
Number of Iterations:  1

Working Correlation
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.6078016 0.6078016
[2,] 0.6078016 1.0000000 0.6078016
[3,] 0.6078016 0.6078016 1.0000000
```

Finally a GEE with exponential correlation structure is fitted:

```
> gee3a <- gee(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit), corstr="AR-M", Mv=1)

> summary(gee3a)

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

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     AR-M , M = 1

Call:
gee(formula = Response ~ Th + Sex + Age + Age2, id = N, family = binomial(link = logit),
    corstr = "AR-M", Mv = 1)

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.8061636 -0.4668263  0.2354196  0.3833613  0.7933803
```

```
Coefficients:
               Estimate  Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -4.72614143 1.912768743 -2.4708379 1.784861526 -2.6479037
Th2         -0.74849866 0.333910055 -2.2416176 0.328084283 -2.2814219
Sex1         0.19277195 0.361954995  0.5325854 0.362465544  0.5318352
Age          0.38489413 0.131285068  2.9317434 0.123289666  3.1218685
Age2        -0.00621548 0.002074949 -2.9954850 0.001945528 -3.1947525

Estimated Scale Parameter:  1.018095
Number of Iterations:  3

Working Correlation
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.7058422 0.4982131
[2,] 0.7058422 1.0000000 0.7058422
[3,] 0.4982131 0.7058422 1.0000000
```

In the following the corresponding marginal models are fitted with the function "geeglm" from the library "geepack".

```
> library(geepack)
```

Model with independent correlation structure:

```
> gee1b <- geeglm(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit))

> summary(gee1b)

Call:
geeglm(formula = Response ~ Th + Sex + Age + Age2, family = binomial(link = logit),
    id = N)

 Coefficients:
             Estimate   Std.err  Wald Pr(>|W|)
(Intercept) -4.749831  1.821209 6.802  0.00911 **
Th2         -0.673979  0.334205 4.067  0.04373 *
Sex1         0.265689  0.366733 0.525  0.46877
Age          0.381151  0.125833 9.175  0.00245 **
Age2        -0.006124  0.001985 9.521  0.00203 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = independence
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    1.006   0.103
Number of clusters:   127  Maximum cluster size: 3
```

Model with exchangeable correlation structure:

```
> gee2b <- geeglm(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit), corstr="exchangeable")

> summary(gee2b)

Call:
geeglm(formula = Response ~ Th + Sex + Age + Age2, family = binomial(link = logit),
    id = N, corstr = "exchangeable")

 Coefficients:
            Estimate  Std.err Wald Pr(>|W|)
(Intercept) -4.74983  1.82121 6.80   0.0091 **
Th2         -0.67398  0.33421 4.07   0.0437 *
Sex1         0.26569  0.36673 0.52   0.4688
Age          0.38115  0.12583 9.17   0.0025 **
Age2        -0.00612  0.00198 9.52   0.0020 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)     1.01   0.103
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.608  0.0883
Number of clusters:   127  Maximum cluster size: 3
```

Model with exponential correlation structure:

```
> gee3b <- geeglm(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit), corstr="ar1")

> summary(gee3b)

Call:
geeglm(formula = Response ~ Th + Sex + Age + Age2, family = binomial(link = logit),
    id = N, corstr = "ar1")

 Coefficients:
            Estimate  Std.err  Wald Pr(>|W|)
(Intercept) -4.72712  1.78605  7.00   0.0081 **
Th2         -0.74443  0.32828  5.14   0.0233 *
Sex1         0.19674  0.36257  0.29   0.5874
Age          0.38467  0.12338  9.72   0.0018 **
Age2        -0.00621  0.00195 10.17   0.0014 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation structure = ar1
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)        1   0.102
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.676  0.0766
Number of clusters:   127  Maximum cluster size: 3
```

For comparison a simple GLM with logit–link is fitted with the same covariates as in the marginal models above:

```
> glm1 <- glm(Response ~ Th + Sex + Age + Age2,
+ family=binomial(link=logit))
> summary(glm1)

Call:
glm(formula = Response ~ Th + Sex + Age + Age2, family = binomial(link = logit))

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.821  -1.139   0.740   0.991   1.705

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.74983    1.26693   -3.75  0.00018 ***
Th2         -0.67398    0.22171   -3.04  0.00237 **
Sex1         0.26569    0.23965    1.11  0.26758
Age          0.38115    0.08674    4.39  1.1e-05 ***
Age2        -0.00612    0.00137   -4.47  7.6e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 513.32  on 380  degrees of freedom
Residual deviance: 478.61  on 376  degrees of freedom
AIC: 488.6

Number of Fisher Scoring iterations: 4
```

It is often advatageous to center the variables like age around a value in the middle of its range. So now the marginal models from above are replicated with age centered around 30 years.

```
> Age <- Age-30
> Age2 <- Age^2
```

Again we use the function "gee" from the "gee"–library for fitting those models.

Model with independent correlation structure and centered age:

```
> gee1c <- gee(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit))

> summary(gee1c)

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

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent

Call:
gee(formula = Response ~ Th + Sex + Age + Age2, id = N, family = binomial(link = logit))

Summary of Residuals:
   Min     1Q Median     3Q    Max
-0.809 -0.477  0.239  0.388  0.766


Coefficients:
            Estimate Naive S.E. Naive z Robust S.E. Robust z
(Intercept)  1.17278    0.28476    4.12     0.44806    2.617
Th2         -0.67398    0.22383   -3.01     0.33421   -2.017
Sex1         0.26569    0.24193    1.10     0.36673    0.724
Age          0.01369    0.01204    1.14     0.01736    0.789
Age2        -0.00612    0.00138   -4.43     0.00198   -3.086

Estimated Scale Parameter:  1.02
Number of Iterations:  1

Working Correlation
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Model with exchangeable correlation structure and centered age:

```
> gee2c <- gee(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit), corstr="exchangeable")

> summary(gee2c)

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

```
Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable

Call:
gee(formula = Response ~ Th + Sex + Age + Age2, id = N, family = binomial(link = logit),
    corstr = "exchangeable")

Summary of Residuals:
   Min     1Q Median     3Q    Max
-0.809 -0.477  0.239  0.388  0.766


Coefficients:
            Estimate Naive S.E. Naive z Robust S.E. Robust z
(Intercept)  1.17278    0.42386   2.767     0.44806    2.617
Th2         -0.67398    0.33316  -2.023     0.33421   -2.017
Sex1         0.26569    0.36012   0.738     0.36673    0.724
Age          0.01369    0.01792   0.764     0.01736    0.789
Age2        -0.00612    0.00206  -2.978     0.00198   -3.086

Estimated Scale Parameter:  1.02
Number of Iterations:  1

Working Correlation
      [,1]  [,2]  [,3]
[1,] 1.000 0.608 0.608
[2,] 0.608 1.000 0.608
[3,] 0.608 0.608 1.000
```

Model with exponential correlation structure and centered age:

```
> gee3c <- gee(Response ~ Th + Sex + Age + Age2, id=N,
+ family=binomial(link=logit), corstr="AR-M", Mv=1)

> summary(gee3c)

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

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     AR-M , M = 1

Call:
gee(formula = Response ~ Th + Sex + Age + Age2, id = N, family = binomial(link = logit),
    corstr = "AR-M", Mv = 1)
```

```
Summary of Residuals:
   Min     1Q Median     3Q    Max
-0.806 -0.467  0.235  0.383  0.793


Coefficients:
            Estimate Naive S.E. Naive z Robust S.E. Robust z
(Intercept)  1.22675    0.42743   2.870     0.44356    2.766
Th2         -0.74850    0.33391  -2.242     0.32808   -2.281
Sex1         0.19277    0.36195   0.533     0.36247    0.532
Age          0.01197    0.01797   0.666     0.01699    0.704
Age2        -0.00622    0.00207  -2.995     0.00195   -3.195

Estimated Scale Parameter:  1.02
Number of Iterations:  3

Working Correlation
      [,1]  [,2]  [,3]
[1,] 1.000 0.706 0.498
[2,] 0.706 1.000 0.706
[3,] 0.498 0.706 1.000
```