Knee Injuries - Proportional Odds Models

February 8, 2012

At the beginning the knee dataset is loaded:

```
> rm(list=ls(all=TRUE))
> library(catdata)
> data(knee)
> attach(knee)
```

First of all a simple χ^2 – test of independence between the therapy (Th) and the pain level (R4) is applied.

```
> suppressWarnings(chisq.test(knee$Th,knee$R4))

        Pearson's Chi-squared test

data:  knee$Th and knee$R4
X-squared = 18.2, df = 4, p-value = 0.001128
```

In the following the variable Age is centered around 30 years and the quadratic variable Age2 is created.

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

The response pain (R4) has to be an ordered factor, the covariates therapy (Th) and sex (Sex) need to be factors.

```
> R4 <- as.ordered(R4)
> Th <- as.factor(Th)
> Sex <- as.factor(Sex)
```

A proportional odds model can be fitted by the function "polr" from the "MASS" – library. Attention has to be paid to the algebraic signs of the coefficients. These are inverse to the usual interpretation in porportional odds models.

```
> library(MASS)
```

The first model only uses therapy as covariate, to achieve a proportional odds model the option "method" needs to use the logistic link function.

```
> polr1 <- polr(R4 ~ Th, method="logistic")

> summary(polr1)
```

```
Call:
polr(formula = R4 ~ Th, method = "logistic")

Coefficients:
     Value Std. Error t value
Th2 -0.893      0.328   -2.72

Intercepts:
    Value  Std. Error t value
1|2 -1.466  0.285     -5.141
2|3 -0.286  0.255     -1.120
3|4  0.667  0.254      2.621
4|5  2.644  0.436      6.059

Residual Deviance: 373.20
AIC: 383.20
```

The corresponding odds–ratio can be recieved by the following command (consider the inverse sign!):

```
> exp(-coef(polr1))
```

```
 Th2
2.44
```

Now a model with the covariates therapy, sex and age is fitted.

```
> polr2 <- polr(R4 ~ Th + Sex + Age, method="logistic")

> summary(polr2)
```

```
Call:
polr(formula = R4 ~ Th + Sex + Age, method = "logistic")

Coefficients:
       Value Std. Error t value
Th2  -0.9438      0.336  -2.813
Sex1  0.0499      0.373   0.134
Age  -0.0159      0.017  -0.936

Intercepts:
    Value  Std. Error t value
1|2 -1.453  0.409     -3.549
2|3 -0.269  0.394     -0.681
3|4  0.686  0.392      1.752
4|5  2.674  0.522      5.121

Residual Deviance: 372.25
AIC: 386.25
```

Odds–ratios for the second model:

```
> exp(-coef(polr2))
```

```
  Th2  Sex1   Age
2.570 0.951 1.016
```

To get the Wald–statistic, the standard errors have to be extracted from the summary. Afterwards the Wald–statistic and the corresponding p–values are easily recieved.

```
> se <- summary(polr2)[1][[1]][1:3,2]
> (wald2 <- -coef(polr2)/se)

   Th2   Sex1    Age
 2.813 -0.134  0.936
```

P–values for the second model:

```
> 1-pchisq(wald2^2, df=1)

    Th2    Sex1     Age
0.00491 0.89367 0.34921
```

Finally the quadratic age–effect is added to the previous model.

```
> polr3 <- update(polr2, ~. + Age2)

> summary(polr3)

Call:
polr(formula = R4 ~ Th + Sex + Age + Age2, method = "logistic")

Coefficients:
        Value Std. Error t value
Th2  -0.94452    0.33871 -2.7886
Sex1 -0.08295    0.37836 -0.2192
Age   0.00171    0.01804  0.0948
Age2 -0.00622    0.00209 -2.9766

Intercepts:
    Value  Std. Error t value
1|2 -2.204  0.490     -4.497
2|3 -0.943  0.460     -2.050
3|4  0.065  0.446      0.145
4|5  2.082  0.557      3.738

Residual Deviance: 362.88
AIC: 378.88
```

Odds–ratios for the final model:

```
> exp(-coef(polr3))

  Th2  Sex1   Age  Age2
2.572 1.086 0.998 1.006
```

Wald–statistic for the final model:

```
> se <- summary(polr3)[1][[1]][1:4,2]
> (wald3 <- -coef(polr3)/se)

    Th2    Sex1     Age    Age2
 2.7886  0.2192 -0.0948  2.9766
```

P–values for the final model:

```
> 1-pchisq(wald3^2, df=1)

    Th2    Sex1     Age    Age2
0.00529 0.82647 0.92445 0.00291
```

As the proportional odds–model is the most popular model for ordinal data, there are several different ways to fit such models. Now the final model is additionally fitted with function "vglm" from the "VGAM"–library and with function "lrm" from the "rms"–library.

Model fitted with "vglm":

```
> library(VGAM)

> m.vglm <- vglm(R4 ~ Th + Sex + Age + Age2, family = cumulative (link="logit",
+ parallel=TRUE))
> summary(m.vglm)

Call:
vglm(formula = R4 ~ Th + Sex + Age + Age2, family = cumulative(link = "logit",
    parallel = TRUE))

Pearson Residuals:
               Min   1Q Median  3Q Max
logit(P[Y<=1])  -2 -0.8   -0.2 0.9 2.9
logit(P[Y<=2])  -3 -0.5    0.3 0.9 1.8
logit(P[Y<=3])  -3 -0.2    0.3 0.4 1.4
logit(P[Y<=4])  -6  0.1    0.1 0.2 0.5

Coefficients:
               Value Std. Error t value
(Intercept):1 -2.204      0.460    -4.8
(Intercept):2 -0.943      0.424    -2.2
(Intercept):3  0.065      0.418     0.2
(Intercept):4  2.082      0.543     3.8
Th2            0.944      0.332     2.8
Sex1           0.083      0.357     0.2
Age           -0.002      0.018    -0.1
Age2           0.006      0.002     3.0

Number of linear predictors:  4
```

```
Names of linear predictors:
logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3]), logit(P[Y<=4])

Dispersion Parameter for cumulative family:   1

Residual Deviance: 363 on 500 degrees of freedom

Log-likelihood: -181 on 500 degrees of freedom

Number of Iterations: 5
```

The resulting coefficients are very similar to the coefficients in the model above fitted with function "polr", but they have inverse signs. Therefore they can be interpreted in the usual way.

Model fitted with "lrm":

```
> library(rms)

> m.lrm <- lrm(R4 ~ Th + Sex + Age + Age2)
> m.lrm

Logistic Regression Model

lrm(formula = R4 ~ Th + Sex + Age + Age2)

Frequencies of Responses

 1  2  3  4  5
36 34 25 26  6

                     Model Likelihood     Discrimination    Rank Discrim.
                        Ratio Test           Indexes           Indexes
Obs           127    LR chi2     17.87    R2        0.138    C       0.656
max |deriv| 3e-10    d.f.            4    g         0.811    Dxy     0.313
                     Pr(> chi2) 0.0013    gr        2.249    gamma   0.317
                                          gp        0.183    tau-a   0.241
                                          Brier     0.210

      Coef    S.E.   Wald Z Pr(>|Z|)
y>=2   2.2040 0.4900  4.50  <0.0001
y>=3   0.9431 0.4600  2.05  0.0403
y>=4  -0.0647 0.4459 -0.15  0.8847
y>=5  -2.0819 0.5568 -3.74  0.0002
Th=2  -0.9445 0.3387 -2.79  0.0053
Sex=1 -0.0829 0.3784 -0.22  0.8265
Age    0.0017 0.0180  0.09  0.9245
Age2  -0.0062 0.0021 -2.98  0.0029
```

Again no big differences are found concerning the coefficients. Here the signs are the same as with the function "polr".