Addiction - Multinomial Model

February 8, 2012

First the "addiction" data are loaded and attached.

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

For the multinomial logit model the function "multinom" from the "nnet"–package is used.

```
> library(nnet)
```

The response "ill" has to be used as factor.

```
> ill <- as.factor(ill)
> addiction$ill<-as.factor(addiction$ill)
```

The first model is a model with the covariates "gender", "university" and a linear effect of "age"

```
> multinom0 <- multinom(ill ~ gender + age + university, data=addiction)
```

```
# weights:  15 (8 variable)
initial  value 749.253581
iter  10 value 675.937605
final  value 675.208456
converged
```

```
> summary(multinom0)
```

```
Call:
multinom(formula = ill ~ gender + age + university, data = addiction)

Coefficients:
  (Intercept)    gender        age university
1   -1.160717 0.4366061 0.02991096   1.622052
2   -2.015571 0.2879080 0.04208660   1.067295

Std. Errors:
  (Intercept)    gender         age university
1   0.2654366 0.1938408 0.006235135  0.2534615
2   0.3076299 0.2207805 0.006821200  0.2891136

Residual Deviance: 1350.417
AIC: 1366.417
```

Another possibility to fit multinomial response models is given by the function "vglm" from the package "VGAM".

```
> library(VGAM)
> multivgam0<-vglm(ill ~ gender + age + university, multinomial(refLevel=1),
+                  data=addiction)
> summary(multivgam0)

Call:
vglm(formula = ill ~ gender + age + university, family = multinomial(refLevel = 1),
    data = addiction)

Pearson Residuals:
                      Min       1Q   Median       3Q    Max
log(mu[,2]/mu[,1]) -4.4464 -0.83311 -0.41954  0.99377 1.5516
log(mu[,3]/mu[,1]) -4.2426 -0.55806 -0.27917 -0.18371 2.4954

Coefficients:
                  Value Std. Error t value
(Intercept):1 -1.160714  0.2654346 -4.3729
(Intercept):2 -2.015564  0.3076272 -6.5520
gender:1       0.436607  0.1938397  2.2524
gender:2       0.287912  0.2207791  1.3041
age:1          0.029911  0.0062350  4.7972
age:2          0.042086  0.0068211  6.1700
university:1   1.622048  0.2534585  6.3997
university:2   1.067287  0.2891095  3.6916

Number of linear predictors:  2

Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])

Dispersion Parameter for multinomial family:   1

Residual Deviance: 1350.417 on 1356 degrees of freedom

Log-likelihood: -675.2085 on 1356 degrees of freedom

Number of Iterations: 4
```

Both models yield the same parameter estimates.

The second model includes an additional quadratic effect of "age".

```
> addiction$age2 <- addiction$age^2
> multinom1 <- update(multinom0, . ~ . + age2)

# weights:  18 (10 variable)
initial  value 749.253581
iter  10 value 666.374546
final  value 658.875161
converged
```

```
> summary(multinom1)

Call:
multinom(formula = ill ~ gender + age + university + age2, data = addiction)

Coefficients:
  (Intercept)    gender       age university         age2
1   -3.720298 0.5264935 0.1840509  1.4546712 -0.001891845
2   -3.502998 0.3562860 0.1357464  0.9362573 -0.001173966

Std. Errors:
  (Intercept)    gender         age university         age2
1 0.011047538 0.1023630 0.008783214 0.11373313 0.0001533591
2 0.008699935 0.0827317 0.009064134 0.09599875 0.0001540031

Residual Deviance: 1317.75
AIC: 1337.75

> multivgam1<-vglm(ill ~ gender + age + university + age2, multinomial(refLevel=1),
+                  data=addiction)
> summary(multivgam1)

Call:
vglm(formula = ill ~ gender + age + university + age2, family = multinomial(refLevel = 1),
    data = addiction)

Pearson Residuals:
                      Min       1Q   Median       3Q    Max
log(mu[,2]/mu[,1]) -3.4647 -0.69123 -0.35630  0.85570 2.7077
log(mu[,3]/mu[,1]) -2.8800 -0.48233 -0.28217 -0.18006 2.8677

Coefficients:
                   Value Std. Error t value
(Intercept):1 -3.7202408 0.54661481 -6.8060
(Intercept):2 -3.5029582 0.59581914 -5.8792
gender:1       0.5264746 0.20083037  2.6215
gender:2       0.3562789 0.22432535  1.5882
age:1          0.1840478 0.02860279  6.4346
age:2          0.1357440 0.03010190  4.5095
university:1   1.4546676 0.25770640  5.6447
university:2   0.9362483 0.29040051  3.2240
age2:1        -0.0018918 0.00033580 -5.6336
age2:2        -0.0011739 0.00033989 -3.4539

Number of linear predictors:  2

Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])

Dispersion Parameter for multinomial family:   1

Residual Deviance: 1317.75 on 1354 degrees of freedom
```

```
Log-likelihood: -658.8752 on 1354 degrees of freedom

Number of Iterations: 4
```

It should be noted that the standard errors for the models generated by "nnet" and "VGAM" differ when age is included quadratically. The parameter estimates are equal again.

Now the necessity of the quadratic term is tested by using the function "anova".

```
> anova(multinom0,multinom1)

Likelihood ratio tests of Multinomial Models

Response: ill
                             Model Resid. df Resid. Dev   Test    Df LR stat.
1        gender + age + university      1356   1350.417
2 gender + age + university + age2      1354   1317.750 1 vs 2     2 32.66659
       Pr(Chi)
1
2 8.063801e-08

> multinom1$dev - multinom0$dev

[1] -32.66659
```

Now we plot the probabilities for the responses against age. First a sequence within the range of age has to be created.

```
> minage <- min(na.omit(age))
> maxage <- max(na.omit(age))
> ageindex <- seq(minage, maxage, 0.1)
> n <- length(ageindex)
```

Now the vectors for the other covariates and the data sets for men and women are built.

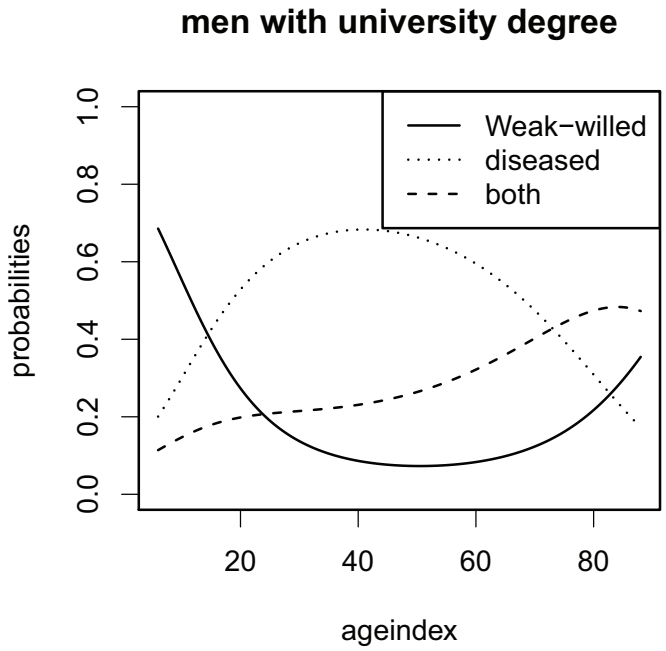
```
> ageindex2 <- ageindex^2
> gender1 <- rep(1, n)
> gender0 <- rep(0, n)
> university1 <- rep(1, n)
> datamale <- as.data.frame(cbind(gender=gender0,age=ageindex,university=
+   university1,age2=ageindex2))
> datafemale <- as.data.frame(cbind(gender=gender1,age=ageindex,university=
+   university1,age2=ageindex2))
```

Now for the built data sets the probabilities based on model "multinom1" are computed.

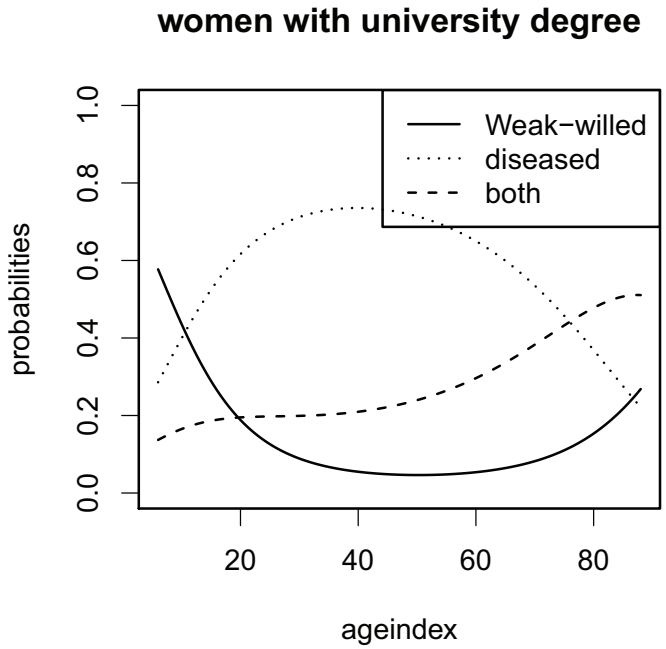
```
> probsmale <- predict(multinom1, datamale, type="probs")
> probsfemale <- predict(multinom1, datafemale, type="probs")
```

Now the probabilities can be plotted.

```
> par(cex=1.4, lwd=2)
> plot(ageindex, probsmale[,1], type="l", lty=1, ylim=c(0,1), main=
+ "men with university degree", ylab="probabilities")
> lines(ageindex, probsmale[,2], lty="dotted")
> lines(ageindex, probsmale[,3], lty="dashed")
> legend("topright", legend=c("Weak-willed", "diseased", "both"), lty=c("solid",
+ "dotted", "dashed"))
```




```
> par(cex=1.4, lwd=2)
> plot(ageindex, probsfemale[,1], type="l", lty=1, ylim=c(0,1), main=
+   "women with university degree", ylab="probabilities")
> lines(ageindex, probsfemale[,2], lty="dotted")
> lines(ageindex, probsfemale[,3], lty="dashed")
> legend("topright", legend=c("Weak-willed", "diseased", "both"),
+        lty=c("solid", "dotted", "dashed"))
```


women with university degree
probabilities
0.0
0.2
0.4
0.6
0.8
1.0
20
40
60
80
ageindex
Weak−willed
diseased
both