

Non-life — Assignment NL3

Niels Keizer* and Robert Jan Sopers†

October 2, 2016

Analyzing a bonus-malus system using GLM

Q1

a)

We are asked to check if the values in Table 9.8 from MART are correct. For this, we first run the code given in the exercise:

```
> rm(list=ls(all=TRUE)) ## First remove traces of previous sessions
> fn <- "http://www1.fee.uva.nl/ke/act/people/kaas/Cars.txt"
> Cars <- read.table(fn, header=TRUE)
> Bminus1 <- Cars$B - 1; Bis14 <- as.numeric(Cars$B==14)
> Cars$A <- as.factor(Cars$A); Cars$R <- as.factor(Cars$R)
> Cars$M <- as.factor(Cars$M); Cars$U <- as.factor(Cars$U)
> Cars$B <- as.factor(Cars$B); Cars$WW <- as.factor(Cars$WW)
> ActualWt <- c(650,750,825,875,925,975,1025,1075,1175,1375,1600)
> W <- log(ActualWt/650)[Cars$WW]
>
> # GLM analysis
>
> g1 <- glm(TotCl/Expo~R+A+U+W+Bminus1+Bis14, quasipoisson, wei=Expo, data=Cars)
> g2 <- glm(TotCl/Expo~R+A+U+W+Bminus1+Bis14+M, quasipoisson, wei=Expo, data=Cars)
> g3 <- glm(TotCl/Expo~R+A+U+W+B, quasipoisson, wei=Expo, data=Cars)
>
> anova(g1,g2)
```

Analysis of Deviance Table

Model 1: TotCl/Expo ~ R + A + U + W + Bminus1 + Bis14

Model 2: TotCl/Expo ~ R + A + U + W + Bminus1 + Bis14 + M

	Resid. Df	Resid. Dev	Df	Deviance
1	7515	38616941		
2	7513	38614965	2	1975.8

```
> anova(g1,g3)
```

*Student number: 10910492

†Student number: 0629049

Analysis of Deviance Table

Model 1: TotCl/Expo ~ R + A + U + W + Bminus1 + Bis14

Model 2: TotCl/Expo ~ R + A + U + W + B

Resid. Df Resid. Dev Df Deviance

1 7515 38616941

2 7504 38544506 11 72435

>

> # Multiplicative coefficients

> options(digits=7)

> exp(coef(g1)); exp(coef(g2)); exp(coef(g3))

(Intercept)	R2	R3	A2	A3	U2	W
524.3016583	1.0842682	1.1916130	0.4147224	0.6184468	1.3841303	2.3722083
Bminus1	Bis14					
0.8978647	1.1053665					
(Intercept)	R2	R3	A2	A3	U2	W
522.6627527	1.0842767	1.1914111	0.4147232	0.6184538	1.3835062	2.3721668
Bminus1	Bis14	M2	M3			
0.8978640	1.1053568	1.0073260	1.0014581			
(Intercept)	R2	R3	A2	A3	U2	W
515.5320549	1.0843018	1.1916593	0.4143437	0.6178700	1.3841612	2.3722369
B2	B3	B4	B5	B6	B7	B8
0.9111279	0.8275175	0.7403718	0.6842609	0.6088526	0.5416103	0.4489065
B9	B10	B11	B12	B13	B14	
0.4151901	0.3888576	0.3459030	0.3143452	0.2832722	0.2773037	

All coefficients can be checked individually against table 9.8 and are the same, except for models g1 and g2, because the bonus malus risk factor is taken as numeric. This means that the factors in the table have been calculated from the factor for *B2* to the power $B - 1$. The coefficient for *Bminus1* differs only in the 7th decimal spot and the table is given with 4 decimals. This means that if we only need to check one of the two models. We do this by recalculating the values in R.

> bm_class <- seq(1,13,1)

> bm_coef <- exp((bm_class-1)*coef(g1)["Bminus1"])

> bm_coef

[1] 1.0000000 0.8978647 0.8061610 0.7238236 0.6498956 0.5835184 0.5239205

[8] 0.4704098 0.4223643 0.3792260 0.3404937 0.3057173 0.2744927

These values also correspond with those in table 9.8

b)

Using the coefficients of g1, g2 and g3, compute the fitted values for the cell 4000.

For this, we use the coefficients in R. Recalculating these by hand would be rather pointless and is an exercise in working neatly over understanding the subject matter.

> # Observed value

```

> g1$y[4000]
4000
326.4545
>
> # Fitted value
> fitted(g1)[4000]; fitted(g2)[4000]; fitted(g3)[4000]
4000
634.0642
4000
636.416
4000
644.5283

```

What we can see is the all three GLM's have a fitted value that is about twice as large as the actual value. Not one of the models is close to the observed value, but the models are quite close together in their estimate.

c)

We now explain the result of the following R-code.

```

> g2$family$linkinv(model.matrix(g2)[4000,]%*%coef(g2))
      [,1]
[1,] 636.416

```

This result is equal to the fitted value of the `g2` model. This is no surprise, considering the code is equal to the definition of the fitted value for cell 4000. The inner product of the values of the risk factors and their corresponding coefficients gives the linear estimator for that cell, after which the `linkinv` function is applied, which is the exponential function. This results in the fitted value.

Q2

First we will determine the scale factor ϕ using a 'rich' model, meaning that the values of both the weight of the car and the BM class are used as factors.

```

> g.rich <- glm(TotCl/Expo~R+A+U+WW+B, quasipoisson, wei=Expo, data=Cars)
> anova(g.rich)
Analysis of Deviance Table

```

Model: quasipoisson, link: log

Response: TotCl/Expo

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
--	----	----------	-----------	------------

NULL			7523	116167018
R	2	2586478	7521	113580540
A	2	23288859	7519	90291681
U	1	4479946	7518	85811735
WW	10	6931993	7508	78879742
B	13	40358336	7495	38521406

We determine the scale factor to be $\frac{38521406}{7495} = 5139.61$. Or in R:

```
> phi <- 38521406/7495
```

To check whether Bis14 can be removed from the model, we use an `anova` call on the model with Bis14 (`g1`) and without (`g.test`).

```
> g.test <- glm(TotCl/Expo~R+A+U+W+Bminus1, quasipoisson, wei=Expo, data=Cars)
> anova(g.test,g1)
Analysis of Deviance Table
```

```
Model 1: TotCl/Expo ~ R + A + U + W + Bminus1
Model 2: TotCl/Expo ~ R + A + U + W + Bminus1 + Bis14
  Resid. Df Resid. Dev Df Deviance
1      7516   38755743
2      7515   38616941  1   138802
```

Next we test if the inclusion of Bis14 is significant:

```
> test <- function (Df, Deviance){
+   scaled.dev <- Deviance/phi
+   test.dev <- qchisq(0.95,Df)
+   return(scaled.dev>test.dev)
+ }
> test(1, 138802)
[1] TRUE
```

First we calculate the scaled deviance. Then we calculate the 95-th percentile of the $\chi^2(k)$ distribution with `Df` degrees of freedom. When the improvement of scaled deviance is larger than the test value, the increase is significant. The test is implemented as a function, so it can be reused in the rest of the exercise. Also the test returns `TRUE`, therefore the inclusion of Bis14 is a significant improvement of the model and can not be removed.

Then we check if B can be removed from model `g3`.

```
> g.test <- glm(TotCl/Expo~R+A+U+W, quasipoisson, wei=Expo, data=Cars)
> anova(g.test,g3)
Analysis of Deviance Table
```

```
Model 1: TotCl/Expo ~ R + A + U + W
Model 2: TotCl/Expo ~ R + A + U + W + B
```

```

  Resid. Df Resid. Dev Df Deviance
1      7517   78902891
2      7504   38544506 13 40358385
> test(13,40358385)
[1] TRUE

```

The test value is TRUE, so B can not be removed from the model. Next we check whether W can be removed from the model:

```

> g.test <- glm(TotCl/Expo~R+A+U+B, quasipoisson, wei=Expo, data=Cars)
> anova(g.test,g3)
Analysis of Deviance Table

Model 1: TotCl/Expo ~ R + A + U + B
Model 2: TotCl/Expo ~ R + A + U + W + B
  Resid. Df Resid. Dev Df Deviance
1      7505   45495122
2      7504   38544506  1  6950616
> test(1,6950616)
[1] TRUE

```

This result implies that W can not be removed from the model g3.

Is it helpful to allow separate coefficients for the weight class in model g1. We again check using the `anova` and `test` functions.

```

> g.test <- glm(TotCl/Expo~R+A+U+WW+Bminus1+Bis14, quasipoisson, wei=Expo, data=Cars)
> anova(g1, g.test)
Analysis of Deviance Table

Model 1: TotCl/Expo ~ R + A + U + W + Bminus1 + Bis14
Model 2: TotCl/Expo ~ R + A + U + WW + Bminus1 + Bis14
  Resid. Df Resid. Dev Df Deviance
1      7515   38616941
2      7506   38593888  9    23053
> test(9,23053)
[1] FALSE

```

This shows that allowing separate coefficients for the weight classes would not be an improvement.

Q3

To answer this question, we run the `test` function defined earlier:

```

> test(7515-7491,38616941-38408588)
[1] TRUE

```

The interaction terms do improve the model significantly. It might be worthwhile to investigate which interaction terms give the most improvement, because there might be some interaction terms which are not significant by themselves.

Q4

First we estimate the number of claims and the size per claim as described. We can combine the two models by adding their coefficients, because directly combining the two models will give a product of two exponentials, which is the same as one exponential with the arguments summed. We compare the resulting coefficients with a direct estimation.

```
> g.nCl <- glm(nCl/Expo~R+A+U+W+Bminus1+Bis14, quasipoisson, wei=Expo, data=Cars)
> g.sCl <- glm(TotCl/nCl~R+A+U+W+Bminus1+Bis14, Gamma(link="log"), wei=nCl, data=Cars)
> g.direct <- glm(TotCl/Expo~R+A+U+W+Bminus1+Bis14, quasipoisson, wei=Expo, data=Cars)
>
> mult.coef <- exp(coef(g.nCl)+coef(g.sCl))
> direct.coef <- exp(coef(g.direct))
> mult.coef; direct.coef
```

(Intercept)	R2	R3	A2	A3	U2
525.1107841	1.0856608	1.1901279	0.4134531	0.6145069	1.3823142
W	Bminus1	Bis14			
2.3827096	0.8979376	1.1057074			

(Intercept)	R2	R3	A2	A3	U2
524.3016583	1.0842682	1.1916130	0.4147224	0.6184468	1.3841303
W	Bminus1	Bis14			
2.3722083	0.8978647	1.1053665			

The resulting models have very similar results and attribute about the same amount of risk to each risk factor.