LogLink Logistic regression AQI data

Michael Andreae

July, 24, 2015

Insurance status predicts antiemetic use

We investigate the Hypothesis that insurance status predicets antiemetic use the population in the Public Use File of the Anestehsia Quality Institute with electronic anesthesia records recording antiemetic use.

Load cleaned dataset *fullAQI_4_14* and *prov1_AQI_4_14*

we load the cleaned larger dataset without and with provider information *fullAQI_4_14.Rdata*, *prov1_AQI_4_14*, which we generated in *import_AQI_14Jul2015.Rmd*

```
rm(list = ls())
load("Analysis/Data/fullAQI_4_14.Rdata")
load("Analysis/Data/prov1_AQI_4_14.Rdata")
str(fullAQI_4_14)
```

```
## 'data.frame':    173133 obs. of  12 variables:
##  $ ond      : Factor w/ 2 levels "no Ondan","Ondan": 2 1 2 1 1 2 2 2 1 2 ...
##  $ dex      : Factor w/ 2 levels "no Dex","Dex": 1 1 1 1 1 1 1 1 1 1 ...
##  $ drop     : Factor w/ 2 levels "no Drope","Drope": 1 1 1 1 1 1 1 1 1 1 ...
##  $ any      : Factor w/ 2 levels "neither","either": 2 1 2 1 1 2 2 2 1 2 ...
##  $ pay      : Factor w/ 4 levels "Commercial","MEDICAID",..: 1 1 1 3 1 3 2 1 3 1 ...
##  $ age      : int  50 53 58 73 64 73 19 27 85 59 ...
##  $ age_group: Factor w/ 6 levels "19 - 49","50 - 64",..: 2 2 2 3 2 3 1 1 4 2 ...
##  $ sex      : Factor w/ 2 levels "female","male": 2 2 1 2 2 1 1 1 2 2 ...
##  $ ASA      : Factor w/ 6 levels "1","2","3","4",..: 2 3 3 2 3 3 2 2 3 2 ...
##  $ duration : int  59 43 190 56 37 116 93 108 70 93 ...
##  $ anes_type: Factor w/ 7 levels "General","Neuroaxial",..: 1 4 1 2 3 1 2 1 1 1 ...
##  $ practice : Factor w/ 4 levels "A","B","C","D": 2 2 2 2 2 2 2 2 2 2 ...
```

```
str(prov1_AQI_4_14)
```

```
## 'data.frame':    146879 obs. of  13 variables:
##  $ ond      : Factor w/ 2 levels "no Ondan","Ondan": 2 1 2 2 2 2 1 1 2 2 ...
##  $ dex      : Factor w/ 2 levels "no Dex","Dex": 1 1 1 1 1 2 1 1 1 2 ...
##  $ drop     : Factor w/ 2 levels "no Drope","Drope": 1 1 1 1 1 1 1 1 1 1 ...
##  $ any      : Factor w/ 2 levels "neither","either": 2 1 2 2 2 2 1 1 2 2 ...
##  $ pay      : Factor w/ 4 levels "Commercial","MEDICAID",..: 3 2 1 1 3 1 1 1 3 1 ...
##  $ age      : int  73 31 56 59 64 49 36 45 62 26 ...
##  $ age_group: Factor w/ 6 levels "19 - 49","50 - 64",..: 3 1 2 2 2 1 1 1 2 1 ...
##  $ sex      : Factor w/ 2 levels "female","male": 1 1 2 2 1 2 2 1 1 2 ...
##  $ ASA      : Factor w/ 6 levels "1","2","3","4",..: 4 3 2 2 3 1 2 2 3 2 ...
##  $ duration : int  172 80 172 122 133 83 136 81 112 110 ...
##  $ anes_type: Factor w/ 7 levels "General","Neuroaxial",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ provider : Factor w/ 720 levels "5622","5623",..: 161 155 152 163 156 156 161 161 156 153 ...
##  $ practice : Factor w/ 4 levels "A","B","C","D": 2 2 2 2 2 2 2 2 2 2 ...
```

Logistic Model 2 Link : log with glm

We try to fit a logistic regression model with the a **log** link with the

formula <- ond ~ pay +age_group +sex +ASA +anes_type +duration +practice

Problem:

We are unable to fit a log link model with all predictors, because we get the error message:

Error: no valid set of coefficients has been found: please supply starting values

if we use more predictors than a few predictors. Two solutions:

- supply starting values for all coefficients, (factors need one for each level)
- consider using the package *biglm*, specifically for log link with large data set

Building Loglink with starting values

Use estimated coefficients as **starting values** for increasingly complex model with one additional predictor per step. Starting values for additional predictors are set to zero.

Estimate fit_log0 with four predictors:

formula0 <- ond ~ pay +age_group +sex

```
formula0 <- ond ~ pay +age_group +sex

fit_log0 <- glm(formula0,
             family = binomial(link = "log"),
             data = fullAQI_4_14)
```

Table 1: Results log0

	OR	p_values
(Intercept)	0.582	0.000
payMEDICAID	0.774	0.000
payMedicare	0.892	0.000
paySELF	1.020	0.368
age_group50 - 64	1.056	0.000
age_group65 - 79	0.935	0.000
age_group80+	0.720	0.000
age_groupUnder 1	0.297	0.000
age_group1-18	1.162	0.000
sexmale	0.961	0.000

Estimate fit_log1 with five predictors:

formula1 <- ond ~ pay +age_group +sex +ASA

```
formula1 <- ond ~ pay +age_group +sex +ASA # one additonal predictor ASA

fit_log1 <- glm(formula1,
             family = binomial(link = "log"),
             data = fullAQI_4_14, start = c(coef(fit_log0),rep(0,5) ))
# start = c( coef(fit_log0), rep(0,5) ) ### For each level of ASA one start value!
```

Table 2: Results log1

	OR	p_values
(Intercept)	0.702	0.000
payMEDICAID	0.838	0.000
payMedicare	0.989	0.113
paySELF	1.080	0.000
age_group50 - 64	1.143	0.000
age_group65 - 79	1.007	0.325
age_group80+	0.819	0.000
age_groupUnder 1	0.337	0.000
age_group1-18	1.070	0.000
sexmale	1.003	0.570
ASA2	0.824	0.000
ASA3	0.803	0.000
ASA4	0.367	0.000
ASA5	0.025	0.000
ASA6	0.032	0.001

Estimate fit_log2 with six predictors:

formula2 <- ond ~ pay +age_group +sex +ASA +anes_type

```
formula2 <- ond ~ pay +age_group +sex +ASA +anes_type

fit_log2 <- glm(formula2,
             family = binomial(link = "log"),
             data = fullAQI_4_14, start = c(coef(fit_log1),rep(0,5) ))
```

Table 3: Results log2

	OR	p_values
(Intercept)	0.891	0.000
payMEDICAID	0.944	0.000
payMedicare	0.971	0.000
paySELF	1.033	0.022
age_group50 - 64	0.970	0.000
age_group65 - 79	0.948	0.000
age_group80+	0.876	0.000
age_groupUnder 1	0.245	0.000
age_group1-18	0.868	0.000
sexmale	0.920	0.000
ASA2	0.980	0.000

	OR	p_values
ASA3	0.880	0.000
ASA4	0.397	0.000
ASA5	0.024	0.000
ASA6	0.029	0.000
anes_typeNeuroaxial	0.311	0.000
anes_typeRegional	0.309	0.000
anes_typeMAC	0.189	0.000
anes_typeSedation	0.123	0.029
anes_typeLocal	0.442	0.332

Estimate fit_log3 with seven predictors:

formula3 <- ond ~ pay +age_group +sex +ASA +anes_type +duration

```
formula3 <- ond ~ pay +age_group +sex +ASA +anes_type +duration

fit_log3 <- glm(formula3,
             family = binomial(link = "log"),
             data = fullAQI_4_14, start = c(coef(fit_log2),rep(0,1) ))
```

Table 4: Results log3

	OR	p_values
(Intercept)	0.887	0.000
payMEDICAID	0.945	0.000
payMedicare	0.972	0.000
paySELF	1.032	0.028
age_group50 - 64	0.970	0.000
age_group65 - 79	0.948	0.000
age_group80+	0.877	0.000
age_groupUnder 1	0.245	0.000
age_group1-18	0.870	0.000
sexmale	0.920	0.000
ASA2	0.980	0.000
ASA3	0.878	0.000
ASA4	0.395	0.000
ASA5	0.024	0.000
ASA6	0.029	0.000
anes_typeNeuroaxial	0.309	0.000
anes_typeRegional	0.309	0.000
anes_typeMAC	0.190	0.000
anes_typeSedation	0.123	0.029
anes_typeLocal	0.444	0.335
duration	1.000	0.013

Estimate fit_log4 with seven predictors:

formula4 <- ond ~ pay +age_group +sex +ASA +anes_type +duration +practice

```
formula4 <- ond ~ pay +age_group +sex +ASA +anes_type +duration +practice

fit_log4 <- glm(formula4,
             family = binomial(link = "log"),
             data = fullAQI_4_14, start = c(coef(fit_log3), rep (0,3) ) )
```

Non convergence *Warning messages: 1: step size truncated due to divergence 2: step size truncated due to divergence 3: step size truncated due to divergence 4: step size truncated due to divergence 5: glm.fit: algorithm did not converge*