Logistic regression AQI data

Michael Andreae

July, 14, 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 Regression

Logistic Model 1 Link: logit

We first fit a logistic regression model with the a **logit** link, saved in *LogisticRegression_AQIFixedEfffects_LogitLink* as shown below:

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

fit_logit <- glm(formula,
                 family = binomial(link = "logit"),
                 data = fullAQI_4_14)
```

Summary: Controlling for age, sex, facilty and case duration, antiemetic administration is strongly associated with insurance status as a marker of SES, but the OR is not as strong as previously estimated controlling for fewer variables. it makes sense that controlling for likely confounders like anesthesia type and ASA status reduces the effect estimate.

Logistic Model 2 Link : log

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

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

Starting values for log link glm

Estimate fit with **one less** predictor:

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

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

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

	OR	X2.5..	X97.5..	p_values
(Intercept)	0.702	0.692	0.711	0.000
payMEDICAID	0.838	0.825	0.850	0.000
payMedicare	0.989	0.977	1.002	0.113
paySELF	1.080	1.038	1.120	0.000
age_group50 - 64	1.143	1.130	1.156	0.000
age_group65 - 79	1.007	0.993	1.022	0.325

	OR	X2.5..	X97.5..	p_values
age_group80+	0.819	0.799	0.839	0.000
age_groupUnder 1	0.337	0.311	0.364	0.000
age_group1-18	1.070	1.052	1.089	0.000
sexmale	1.003	0.994	1.012	0.570
ASA2	0.824	0.813	0.836	0.000
ASA3	0.803	0.791	0.815	0.000
ASA4	0.367	0.358	0.376	0.000
ASA5	0.025	0.015	0.038	0.000
ASA6	0.032	0.002	0.137	0.001