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 *myAQI_4_14.Rdata*

we load the cleaned larger dataset without provider information *fullAQI_4_14.Rdata*, 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 fit a logistic regression model with the a **logit** link.

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

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

	OR	p_values
(Intercept)	5.796	0.000
payMEDICAID	0.804	0.000
payMedicare	0.904	0.000
paySELF	0.825	0.002
age_group50 - 64	0.886	0.000
age_group65 - 79	0.807	0.000
age_group80+	0.692	0.000
age_groupUnder 1	0.078	0.000
age_group1-18	0.640	0.000
sexmale	0.735	0.000
ASA2	0.875	0.000
ASA3	0.571	0.000
ASA4	0.119	0.000
ASA5	0.005	0.000
ASA6	0.004	0.000
anes_typeNeuroaxial	0.090	0.000
anes_typeRegional	0.103	0.000
anes_typeMAC	0.053	0.000
anes_typeSedation	0.020	0.000
anes_typeLocal	0.195	0.163
duration	0.999	0.000
practiceB	1.270	0.000
practiceC	2.444	0.000
practiceD	1.123	0.000

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.

```
# formula <- ond ~ pay +age_group +sex +ASA +anes_type +duration +practice
#
# fit_logistic <- glm(formula,
```

```
#                     family = binomial(link = "log"),
#                     data = fullAQI_4_14)
```

But we get this error message: *Error: no valid set of coefficients has been found: please supply starting values*

We do not get an error using fewer predictors!

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

fit_logistic <- glm(formula,
                    family = binomial(link = "log"),
                    data = fullAQI_4_14)
```

	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

Logistig Model 3 with random effects provider

We try to add **providers** as random effects and fit a logistic regression model with the a logit link again using the package *lme4*

Only a very simple model converges so far:

```
formula <- ond ~  pay  +(1|provider)

fit_random <- glmer(formula,
                    data = prov1_AQI_4_14,
                    family = binomial,
                    control = glmerControl(optimizer = "bobyqa"),
                    nAGQ = 10)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 10) [glmerMod]
##  Family: binomial  ( logit )
## Formula: ond ~ pay + (1 | provider)
##    Data: prov1_AQI_4_14
##       AIC       BIC    logLik  deviance  df.resid
## 184941.48 184990.97 -92465.74 184931.48    146874
## Random effects:
##  Groups   Name        Std.Dev.
##  provider (Intercept) 0.9613
## Number of obs: 146879, groups:  provider, 720
```

```
## Fixed Effects:
## (Intercept)  payMEDICAID  payMedicare      paySELF
##      0.2650      -0.4646      -0.5605      -0.1352
```

The more complex random effects model fails to converge:

```
formula <- ond ~
  pay +age_group +sex +ASA +anes_type +duration +practice +(1|provider)

# fit_random <- glmer(formula,
#                     data = prov1_AQI_4_14,
#                     family = binomial,
#                     control = glmerControl(optimizer = "bobyqa"),
#                     nAGQ = 10)
```

Error message: *Warning messages: 1: In checkConv(attr(opt, "derivs"), optpar, ctrl = control checkConv, : unable to evaluate scaled gradient 2: In checkConv(attr(opt, "derivs"), optpar, ctrl = control checkConv, : Model failed to converge: degenerate Hessian with 1 negative eigenvalues*