

Homework 3

STAT 340: Applied Regression

Data were collected on doctor visits from a sample of 5,190 people in the 1977/1978 Australian Health Survey. Cameron and Trivedi (1986) sought to explain the variation in doctor visits using one or more explanatory variables. The data can be found in an R data set from `library(AER)` accessible with the command `data("DoctorVisits")`. Variable descriptions can be found under `help("DoctorVisits")`.

Zero-inflated Poisson

```
## Read in data from AER package
data("DoctorVisits")
head(DoctorVisits)
```

```
##   visits gender  age income illness reduced health private freepoor freerepat
## 1      1 female 0.19  0.55        1        4        1     yes        no        no
## 2      1 female 0.19  0.45        1        2        1     yes        no        no
## 3      1  male 0.19  0.90        3        0        0     no        no        no
## 4      1  male 0.19  0.15        1        0        0     no        no        no
## 5      1  male 0.19  0.45        2        5        1     no        no        no
## 6      1 female 0.19  0.35        5        1        9     no        no        no
##   nchronic lchronic
## 1       no       no
## 2       no       no
## 3       no       no
## 4       no       no
## 5      yes       no
## 6      yes       no
```

Investigate the distribution of visits graphically. What do you observe about this distribution?

Consider the following zero-inflated Poisson (ZIP) model:

```
zip1 <- zeroinfl(visits ~ illness*age + reduced*gender | private + freepoor + freerepat,
                 data = DoctorVisits)
summary(zip1)
```

```
##
## Call:
## zeroinfl(formula = visits ~ illness * age + reduced * gender | private +
##         freepoor + freerepat, data = DoctorVisits)
```

```
##
## Pearson residuals:
##      Min      1Q  Median      3Q      Max
## -1.3567 -0.4716 -0.3776 -0.2930 11.4124
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.901970   0.132142 -14.393 < 2e-16 ***
## illness       0.335202   0.047443   7.065 1.60e-12 ***
## age           0.651747   0.241417   2.700 0.00694 **
## reduced       0.148159   0.007916  18.718 < 2e-16 ***
## genderfemale  0.288452   0.070631   4.084 4.43e-05 ***
## illness:age   -0.247367   0.086591  -2.857 0.00428 **
## reduced:genderfemale -0.025454  0.009497  -2.680 0.00736 **
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1521     0.1620  -0.939 0.3478
## privateyes   -0.5486     0.1799  -3.049 0.0023 **
## freepooryes  0.8900     0.3372   2.639 0.0083 **
## freerepatyes -1.1834     0.2533  -4.673 2.97e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 19
## Log-likelihood: -3270 on 11 Df
```

```
exp(confint(zip1))
```

```
##              2.5 %    97.5 %
## count_(Intercept) 0.1152141 0.1934034
## count_illness      1.2740687 1.5344753
## count_age          1.1955162 3.0799597
## count_reduced      1.1418445 1.1778294
## count_genderfemale 1.1618547 1.5324793
## count_illness:age  0.6589663 0.9252871
## count_reduced:genderfemale 0.9568892 0.9931838
## zero_(Intercept)  0.6252112 1.1799308
## zero_privateyes    0.4060544 0.8220644
## zero_freepooryes   1.2575039 4.7157046
## zero_freerepatyes  0.1864117 0.5030480
```

Write out the form of the model fit in zip1. Be sure to include both the latent logistic regression and Poisson regression pieces of this model, and define the notation you use to express your model.

Interpret the 95% confidence interval for the effect of having free government health insurance due to low income (freepooryes) on the odds of not going to the doctor.

Consider the 95% confidence intervals for the effects of having private health insurance or free government health insurance due to old age, disability, or veteran status. How do these effects compare to the effect of having free government health insurance due to low income? What is a reasonable explanation for the differences in these effects?

How does the effect of the number of days of reduced activity in the past 2 weeks due to illness or injury (reduced) on the mean number of doctor visits in the past 2 weeks differ for those recorded as male versus female?