Chapter 4 Guided 6

Professor George

May 4, 2021

A

```
library(AER)
library(ggplot2)
library(gridExtra)
library(tidyverse)
library(pscl)
data('NMES1988')
NMES = NMES1988
names(NMES)
```

```
##  [1] "visits"    "nvisits"   "ovisits"   "novisits"  "emergency" "hospital"
##  [7] "health"    "chronic"   "adl"       "region"    "age"       "afam"
## [13] "gender"    "married"   "school"    "income"    "employed"  "insurance"
## [19] "medicaid"
```

```
# Create a histogram of the visits
obsBar = ggplot(NMES,aes(x=visits))+
  geom_bar()

# Create a histogram of the poisson distribution using the mean
# of visits and the max of visits as a visual maximum
sum1<-NMES%>%
  summarise(lambda=mean(visits),
            maxvisits = max(visits))
sum1
```

```
##     lambda maxvisits
## 1 5.774399        89
```

```
possible.values = with(sum1,0:maxvisits)
possible.values
```

```
##  [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
## [26] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49
## [51] 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74
## [76] 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89
```

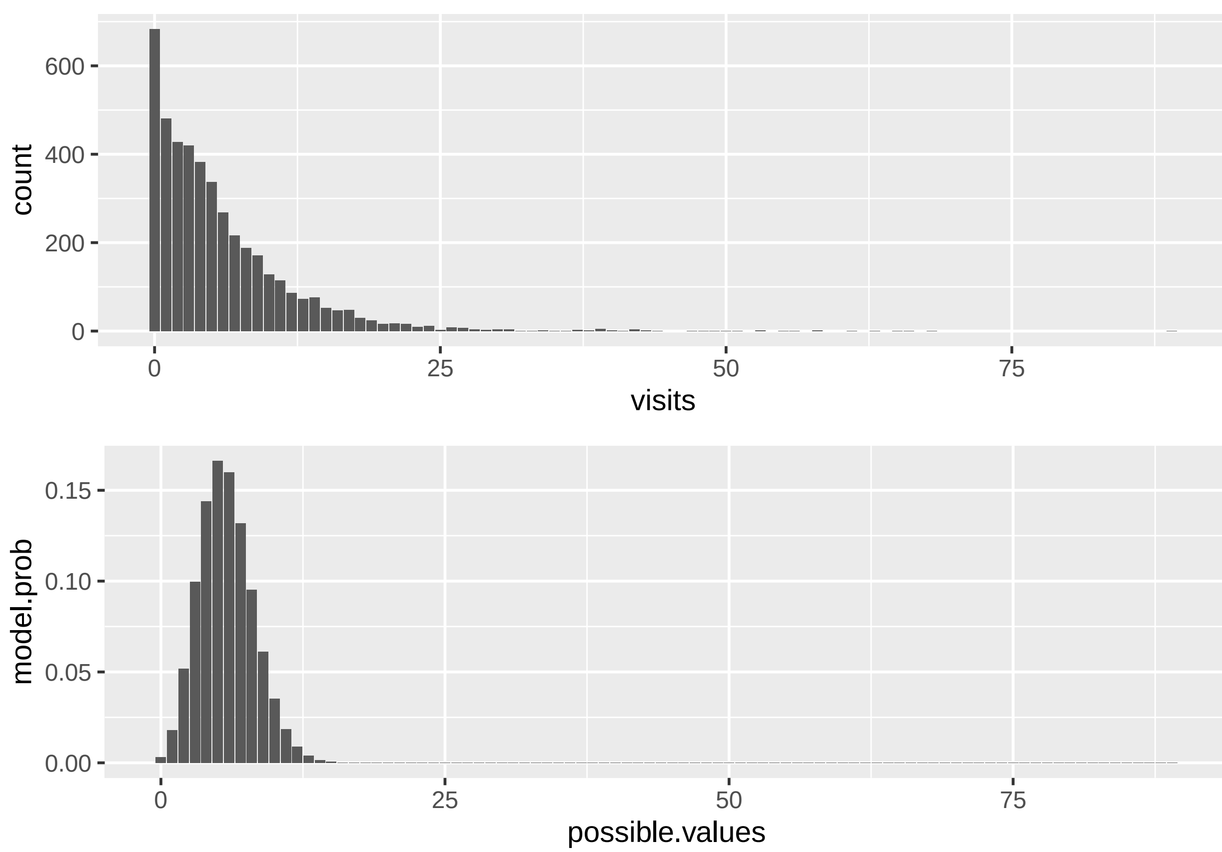
```
model.prob <- with(sum1,dpois(possible.values,
                              lambda ))
# Look at the first few probabilities to see if they look
# realistic (0 to 1)
model.prob[1:10]
```

```
##  [1] 0.003106065 0.017935659 0.051783821 0.099673473 0.143888589 0.166174012
##  [7] 0.159925829 0.131925068 0.095223490 0.061095376
```

```
pois.model = data.frame(possible.values,
                        model.prob)

# Make the 2nd histogram of the poisson
expData = ggplot(pois.model,aes(y=model.prob,
       x=possible.values))+
  geom_bar(stat='identity')

grid.arrange(obsBar,expData)
```



We can see that the two distributions do not match!

B

Fitting the indicated ZIP model

```r
# Terms on right side of | are those we are going to use
# for a logistic model where we are predicting the probability
# that an older adult does *not* go to a physician.
model1 <- zeroinfl(visits~chronic+health+insurance|
                chronic+insurance,data=NMES)
summary(model1)
```

```
##
## Call:
## zeroinfl(formula = visits ~ chronic + health + insurance | chronic +
##     insurance, data = NMES)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -3.9221 -1.2195 -0.4316  0.5598 24.1031
##
## Count model coefficients (poisson with log link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.55878    0.01762  88.448   <2e-16 ***
## chronic          0.11868    0.00462  25.691   <2e-16 ***
## healthpoor       0.29470    0.01729  17.043   <2e-16 ***
## healthexcellent -0.30482    0.03115  -9.786   <2e-16 ***
## insuranceyes     0.14467    0.01631   8.870   <2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.37426    0.09213  -4.062 4.86e-05 ***
## chronic      -0.56112    0.04334 -12.948  < 2e-16 ***
## insuranceyes -0.88314    0.09464  -9.332  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 11
## Log-likelihood: -1.651e+04 on 8 Df
```

```r
coef(summary(model1))
```

```
## $count
##                   Estimate  Std. Error   z value      Pr(>|z|)
## (Intercept)      1.5587819 0.017623698 88.448056  0.000000e+00
## chronic          0.1186784 0.004619517 25.690656 1.486613e-145
## healthpoor       0.2947028 0.017291255 17.043459  3.909129e-65
## healthexcellent -0.3048182 0.031147632 -9.786239  1.290055e-22
## insuranceyes     0.1446700 0.016309360  8.870367  7.290531e-19
##
## $zero
##                Estimate Std. Error    z value     Pr(>|z|)
## (Intercept)  -0.3742561 0.09212590  -4.062441 4.856218e-05
## chronic      -0.5611181 0.04333785 -12.947530 2.426374e-38
## insuranceyes -0.8831440 0.09463925  -9.331689 1.041975e-20
```

```
exp(coef(summary(model1))$count)
```

```
##                  Estimate Std. Error      z value Pr(>|z|)
## (Intercept)     4.7530279   1.017780 2.585251e+38        1
## chronic         1.1260078   1.004630 1.436515e+11        1
## healthpoor      1.3427272   1.017442 2.522784e+07        1
## healthexcellent 0.7372574   1.031638 5.621995e-05        1
## insuranceyes    1.1556581   1.016443 7.117892e+03        1
```

```
exp(coef(summary(model1))$zero)
```

```
##               Estimate Std. Error      z value Pr(>|z|)
## (Intercept)  0.6878008   1.096503 1.720696e-02 1.000049
## chronic      0.5705708   1.044291 2.382097e-06 1.000000
## insuranceyes 0.4134809   1.099262 8.857256e-05 1.000000
```

Interpretations

Poisson Part of Model

Chronic

After exp, 1.126

For a unit increase in the number of chronic conditions a older adult inclined to visit physicians has, holding health status and their status of private insurance constant, we expect an increase on average of 12.6% more visits to a physician.

Poor Health

After exponentiation, 1.34

We expect older adults who visit physicians to visit physicians 34% more on average if they have poor health vs average health, holding private insurance status and number of chronic conditions constant.

Logistic Part of Model

Intercept

After exponentiation, .688

For adults with 0 chronic conditions and no private health insurance, the odds of them *not* visiting a physician is about .688.

Thus the estimate of the probability they do not visit a physician is $\frac{.68}{1+.68} = .404$. And the probability they do visit a physician is 1-.404 = .596.

Insurance

After exponentiation with reference level as a "no", .413

This represents the change in the odds of someone not going to a physician when they have private insurance vs when they do not.

The odds that an older adult will not go to a physician when they have private insurance is .413 of the odds than when they do have insurance, holding chronic illnesses constant.

Note that it is *not* correct to do 1-.41 and reverse this statement since this is change in odds and not a probability.

This is very annoying/difficult to interpret. Let us instead recode our insurance variable so 1 is yes and 2 is no. Then refit.

```
# In order to get a coefficient of insurance that is easier to
# interpret, we can reorder that varible (change the regression
# reference level)
NMES$insurance = factor(NMES$insurance,levels = c("yes","no"))
str(NMES)
```

```
## 'data.frame':    4406 obs. of  19 variables:
##  $ visits   : int  5 1 13 16 3 17 9 3 1 0 ...
##  $ nvisits  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ovisits  : int  0 2 0 5 0 0 0 0 0 0 ...
##  $ novisits : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ emergency: int  0 2 3 1 0 0 0 0 0 0 ...
##  $ hospital : int  1 0 3 1 0 0 0 0 0 0 ...
##  $ health   : Factor w/ 3 levels "poor","average",..: 2 2 1 1 2 1 2 2 2 2 ...
##   ..- attr(*, "contrasts")= num [1:3, 1:2] 1 0 0 0 0 1
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:3] "poor" "average" "excellent"
##   .. .. ..$ : chr [1:2] "poor" "excellent"
##  $ chronic  : int  2 2 4 2 2 5 0 0 0 0 ...
##  $ adl      : Factor w/ 2 levels "normal","limited": 1 1 2 2 2 2 1 1 1 1 ...
##  $ region   : Factor w/ 4 levels "northeast","midwest",..: 4 4 4 4 4 4 2 2 2 2 ...
##   ..- attr(*, "contrasts")= num [1:4, 1:3] 1 0 0 0 0 1 0 0 0 0 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:4] "northeast" "midwest" "west" "other"
##   .. .. ..$ : chr [1:3] "northeast" "midwest" "west"
##  $ age      : num  6.9 7.4 6.6 7.6 7.9 6.6 7.5 8.7 7.3 7.8 ...
##  $ afam     : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 1 1 1 ...
##  $ gender   : Factor w/ 2 levels "female","male": 2 1 1 2 1 1 1 1 1 1 ...
##  $ married  : Factor w/ 2 levels "no","yes": 2 2 1 2 2 1 1 1 1 1 ...
##  $ school   : int  6 10 10 3 6 7 8 8 8 8 ...
##  $ income   : num  2.881 2.748 0.653 0.659 0.659 ...
##  $ employed : Factor w/ 2 levels "no","yes": 2 1 1 1 1 1 1 1 1 1 ...
##  $ insurance: Factor w/ 2 levels "yes","no": 1 1 2 1 1 2 1 1 1 1 ...
##  $ medicaid : Factor w/ 2 levels "no","yes": 1 1 2 1 1 2 1 1 1 1 ...
```

```
head(NMES)
```

```
##   visits nvisits ovisits novisits emergency hospital  health chronic     adl
## 1      5       0       0        0         0        1 average       2  normal
```

```
## 2      1       0       2        0         2        0 average       2  normal
## 3     13       0       0        0         3        3    poor       4 limited
## 4     16       0       5        0         1        1    poor       2 limited
## 5      3       0       0        0         0        0 average       2 limited
## 6     17       0       0        0         0        0    poor       5 limited
##   region age afam gender married school income employed insurance medicaid
## 1  other 6.9  yes   male     yes      6 2.8810      yes       yes       no
## 2  other 7.4   no female     yes     10 2.7478       no       yes       no
## 3  other 6.6  yes female      no     10 0.6532       no        no      yes
## 4  other 7.6   no   male     yes      3 0.6588       no       yes       no
## 5  other 7.9   no female     yes      6 0.6588       no       yes       no
## 6  other 6.6   no female      no      7 0.3301       no        no      yes
```

```r
# Refint the model using the data after the reorder
model2 <- zeroinfl(visits~chronic+health+insurance|
                chronic+insurance,data=NMES)
summary(model2)
```

```
##
## Call:
## zeroinfl(formula = visits ~ chronic + health + insurance | chronic +
##     insurance, data = NMES)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -3.9221 -1.2195 -0.4316  0.5598 24.1031
##
## Count model coefficients (poisson with log link):
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.70345    0.01107 153.911   <2e-16 ***
## chronic          0.11868    0.00462  25.691   <2e-16 ***
## healthpoor       0.29470    0.01729  17.043   <2e-16 ***
## healthexcellent -0.30482    0.03115  -9.786   <2e-16 ***
## insuranceno     -0.14467    0.01631  -8.870   <2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.25740    0.06722 -18.705   <2e-16 ***
## chronic     -0.56112    0.04334 -12.948   <2e-16 ***
## insuranceno  0.88314    0.09464   9.332   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 11
## Log-likelihood: -1.651e+04 on 8 Df
```

```r
coef(summary(model2))
```

```
## $count
##                   Estimate  Std. Error    z value      Pr(>|z|)
## (Intercept)      1.7034519 0.011067791 153.910736  0.000000e+00
## chronic          0.1186784 0.004619517  25.690655 1.486661e-145
## healthpoor       0.2947028 0.017291255  17.043459  3.909145e-65
```

```
## healthexcellent -0.3048182 0.031147632  -9.786239  1.290053e-22
## insuranceno     -0.1446700 0.016309360  -8.870367  7.290495e-19
##
## $zero
##                Estimate Std. Error    z value     Pr(>|z|)
## (Intercept) -1.2574000 0.06722219 -18.705134 4.495986e-78
## chronic     -0.5611182 0.04333785 -12.947531 2.426332e-38
## insuranceno  0.8831440 0.09463925   9.331688 1.041976e-20
```

```
exp(coef(summary(model2))$count)
```

```
##                  Estimate Std. Error      z value Pr(>|z|)
## (Intercept)     5.4928755   1.011129 6.959588e+66        1
## chronic         1.1260078   1.004630 1.436513e+11        1
## healthpoor      1.3427272   1.017442 2.522784e+07        1
## healthexcellent 0.7372574   1.031638 5.621994e-05        1
## insuranceno     0.8653078   1.016443 1.404910e-04        1
```

```
exp(coef(summary(model2))$zero)
```

```
##              Estimate Std. Error      z value Pr(>|z|)
## (Intercept) 0.2843925   1.069533 7.524253e-09        1
## chronic     0.5705707   1.044291 2.382093e-06        1
## insuranceno 2.4184916   1.099262 1.129018e+04        1
```

Now let us interpret the coefficient for insurance.

Holding chronic illnesses constant, the odds that someone without insurance not visiting a physician is 2.42 higher than a person with private insurance.

C

```
# Fit a standard poisson model
model2 <- glm(visits ~ insurance + chronic
                + health,data=NMES,family=poisson)
summary(model2)
```

```
##
## Call:
## glm(formula = visits ~ insurance + chronic + health, family = poisson,
##     data = NMES)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.2044  -1.9838  -0.7370   0.7142  16.1283
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.49598    0.01081  138.44   <2e-16 ***
```

```
## insuranceno     -0.27862    0.01613  -17.27   <2e-16 ***
## chronic          0.16728    0.00446   37.50   <2e-16 ***
## healthpoor       0.29043    0.01734   16.75   <2e-16 ***
## healthexcellent -0.35961    0.03025  -11.89   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 26943  on 4405  degrees of freedom
## Residual deviance: 24076  on 4401  degrees of freedom
## AIC: 36862
##
## Number of Fisher Scoring iterations: 5
```

```
# Conduct the voung test
vuong(model2,model1)
```

```
## Vuong Non-Nested Hypothesis Test-Statistic:
## (test-statistic is asymptotically distributed N(0,1) under the
##  null that the models are indistinguishible)
## -------------------------------------------------------------
##               Vuong z-statistic             H_A    p-value
## Raw                   -18.18176 model2 > model1 < 2.22e-16
## AIC-corrected         -18.15321 model2 > model1 < 2.22e-16
## BIC-corrected         -18.06201 model2 > model1 < 2.22e-16
```

In all tests of the vuong test, the null is rejected. Our zero inflated poisson model is significantly better than our basic poisson model.