Midterm #2 Question 2 makeup

Luke Letizia

May 3, 2018

Question 2

Introduction

The car insurance company has already implemented a poisson regression model to describe the claim counts distribution. I have been hired to test whether the poisson model should be replaced by a homogeneous negative binomial model. I have concluded that a Negative Binomial model better fits the data and should be used to describe the claim counts distribution.

Methods

Linders Consulting is implementing a poisson regression model to describe the claim counts distribution. Because λ is the parameter of a poisson distribution and is the value for both the mean and variance, it may be inaccurate to assume all samples follow the same λ variable given different people have different behaviors and different risk factors. When the mean and variance are not equal in a counts distribution, there is an under or overdispersion, depending on if the variance is smaller or larger than the mean. The issue could be resolved by implementing a homogeneous negative binomial regression model. The variance is $\lambda(1 + \lambda/\gamma)$ and mean of λ. After implementing the negative binomial and finding the maximum predicted log likelihood λ and γ, I used the persp plotting function to prove the predicted values were the absolute maximums. Another method used to conclude the negative binomial was comparing the chi-square statistics of the negative binomial and the possion regression.

Results

By comparing the poisson regression and the negative binomial, I have concluded that the negative binomial is favored over the poisson because the predicted values are much closer and more honed in to the actual data set than the poisson model is. The chi-square statistic for the negative binomial is also much closer to 0 which proves it is superior.

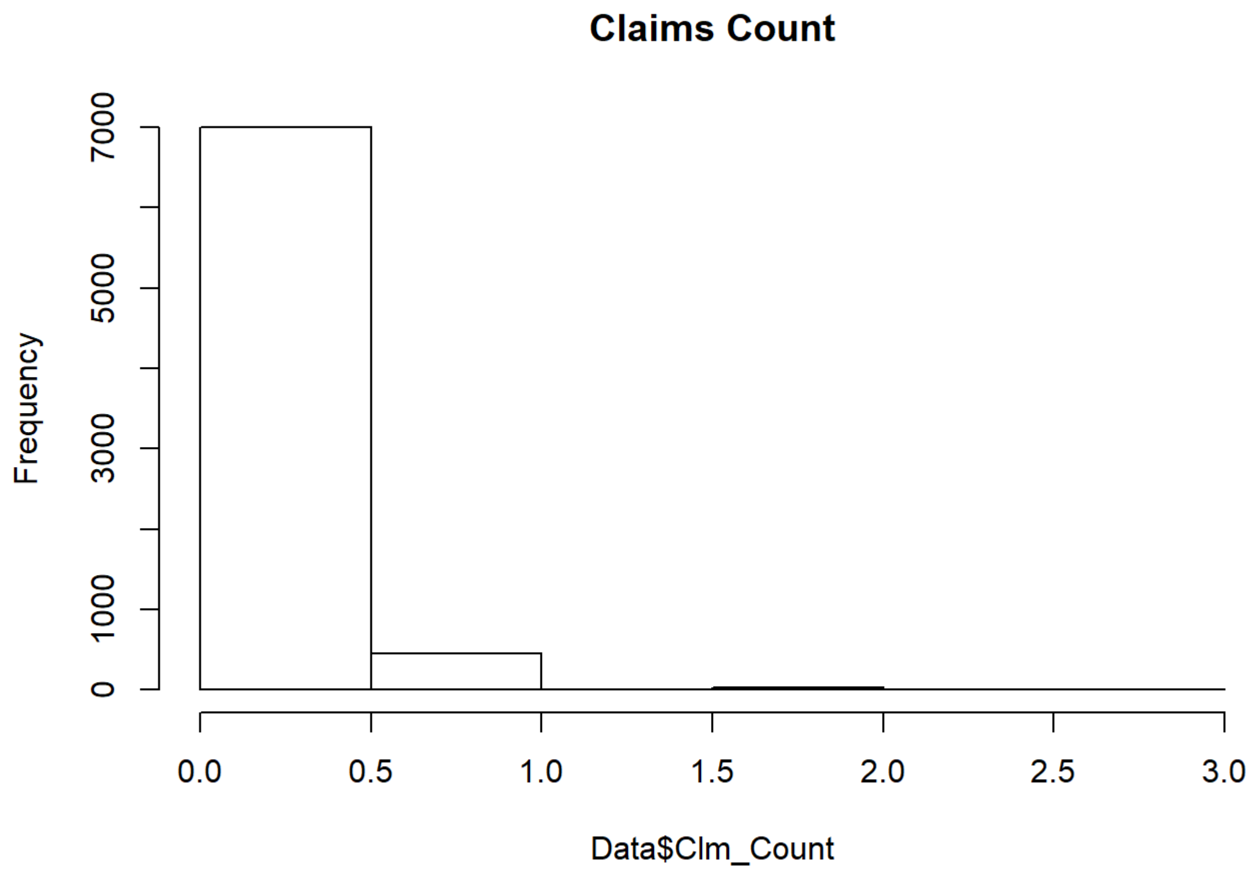
Conclusion

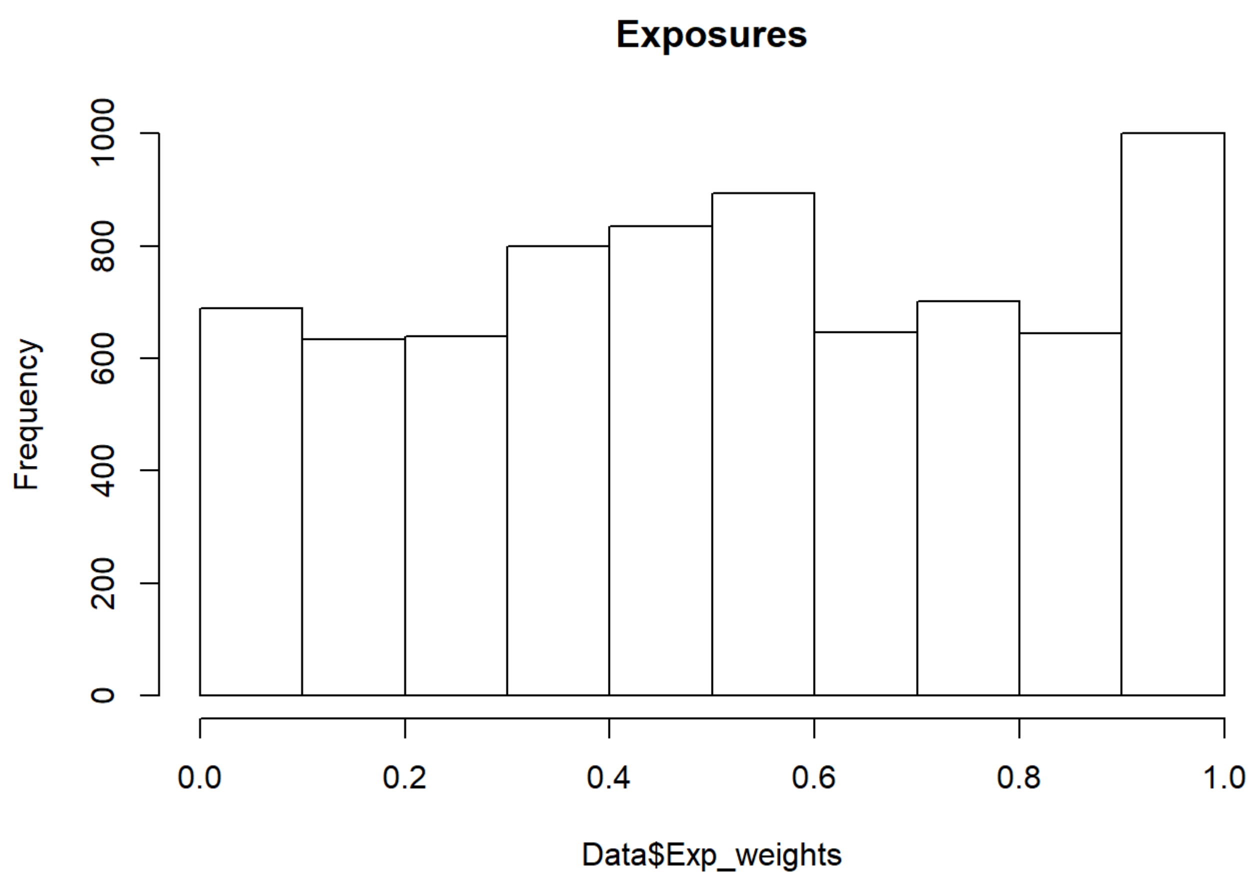
Through the methods used and based on the statistical proof from the dataset provided, I conlcude that the homogeneous negative should replace the poisson regression model. By using the negative binomial, future premiums should thus decrease becuase it increases the predicted number that are accident free and decreases the number that have one or more accidents.

Implementations

The summary of the data doesn’t really give us much information to go off of for the claim counts so we look at the histogram to get a better visualization of the data.

Processing math: 100%





Processing math: 100%

```
##  #Claims Observed
##        0     6996
##        1      455
##        2       28
##        3        4
```

Implementing a Homogeneous Poisson Model

```
##  #Claims    Data Model without Exposure Model With Exposure
##     0.00 6996.00                 6977.86            6983.05
##     1.00  455.00                  487.69             477.67
##     2.00   28.00                   17.04              21.52
##     3.00    4.00                    0.40               0.73
```

We can see that the model without exposure differs from the actual data observed. We can see that including exposures is necesarry for the model because it hones the values for each claim count closer to the observed values, but it is still not an adeqaute fit.

Chi Square Statistic without exposures:

```
## [1] 42
```

Chi Square Statistic with exposures:

```
## [1] 18
```

Because the Chi-Square Satistic is much lower with exposures, we can see that Exposures are an important factor to include when modeling the data.

Generalized Linear Poisson Model

We now investigate modeling different risk parameters for different drivers based on factors involved.

Comparing the Homogeneous Poisson and Poisson Regression

```
##  #Claims    Data Model without Exposure Model With Exposure
##     0.00 6996.00                 6977.86            6983.05
##     1.00  455.00                  487.69             477.67
##     2.00   28.00                   17.04              21.52
##     3.00    4.00                    0.40               0.73
##  Poisson Regression
##             6986.94
##              470.30
##               24.63
##                1.09
```

Fit Negative Binomial

Processing math: 100%

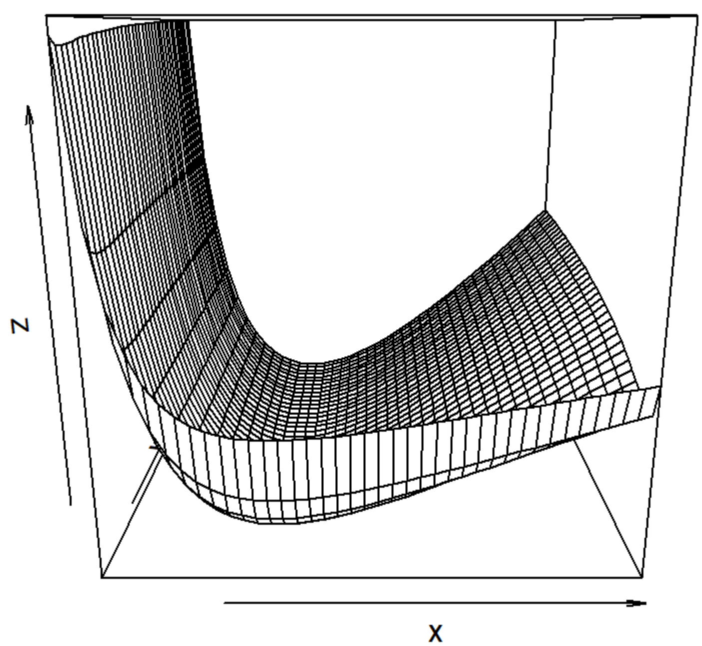
```
## Warning in log((gamma(N + gam))/(gamma(gam) * factorial(N)) * ((1 - ((lamb
## * : NaNs produced

## Warning in log((gamma(N + gam))/(gamma(gam) * factorial(N)) * ((1 - ((lamb
## * : NaNs produced

## Warning in log((gamma(N + gam))/(gamma(gam) * factorial(N)) * ((1 - ((lamb
## * : NaNs produced
```

Max λ = 0.13 Max γ = 1.42

Plotted Log Likelihood Function for different values of λ and γ



Comparing the Negative Binomial and the Poisson Regression

```
##  #Claims   Data Negative Binomial
##      0.0 6996.0            6997.3
##      1.0  455.0             451.0
##      2.0   28.0              32.1
##      3.0    4.0               2.4
```

Processing math: 100%

```
##  #Claims   Data Poisson Regression Negative Binomial
##      0.0 6996.0             6986.9            6997.3
##      1.0  455.0              470.3             451.0
##      2.0   28.0               24.6              32.1
##      3.0    4.0                1.1               2.4
```

Negative Binomial Chi-Square Statistic

```
## [1] 1.6
```

Poisson Regression Chi-Square Statistic

```
## [1] 8.7
```

Proofs

$E[N] = E[E[N \mid \theta]] = E[\theta\lambda] = \lambda E[\theta] = \lambda * 1 = \lambda$

$Var[N] = E[Var[N \mid \theta]] + Var[E[N \mid \theta]] = E[Var[\theta\lambda]] + Var[E[[\theta\lambda]] = \lambda^2 1/\gamma + Var[\lambda] = \lambda^2 1/\gamma + \lambda = \lambda + (1 + \lambda/\gamma)$

Overdispersion is when there is a presence of a greater variability than should be expected. As we can see with the Negative Binomial, it has a smaller dispersion and thus is favored over Poisson.

Processing math: 100%