

Data Analysis for Motor insurance Data

"GLM model using R software"

Ahmed Sabeel
sabeel2010@gmail.com

June 20, 2020

```
## Warning: package 'xtable' was built under R version 3.6.3
## Warning: package 'car' was built under R version 3.6.3
## Loading required package: carData
## Warning: package 'carData' was built under R version 3.6.1
##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary
Statistics Tables.
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
## Warning: package 'ggplot2' was built under R version 3.6.3
## Installing package into 'C:/Users/USER/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
## Error in contrib.url(repos, "source"): trying to use CRAN without setting
a mirror
## Warning: package 'writexl' was built under R version 3.6.3
```

Descriptive Statistics

Read the data from the motor insurance data into R file from Excel and SPSS data files

```
## re-encoding from CP1252

##                25%
## Mean           0.4511
## Standard Deviation 0.7784
## Minimum        0.0000
## First Quantile  0.0000
## Median         0.0000
## Third Quantile  1.0000
```

```
## 90% Quantile      1.0000
## 95% Quantile      2.0000
## 99% Quantile      3.0000
## Maximum           8.0000
##
##                      25%
## Mean              3439.958
## Standard Deviation 14766.744
## Minimum           0.000
## First Quantile     0.000
## Median             0.000
## Third Quantile     934.000
## 90% Quantile       8668.000
## 95% Quantile       16050.800
## 99% Quantile       52610.341
## Maximum            450000.000
```

Frequency model of vehicle data

Performing GLM Calcs Using Poisson, negative Binomial and Quasi Poisson Distributions

```
##
## Call:
## glm(formula = q11 ~ q1 + q2 + q3 + q4 + q5 + q6 + q7 + q8 + q9 +
##      q10 + offset(log(q13)), family = poisson(link = log), data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4013  -0.9830  -0.8757   0.6023   5.3780
##
## Coefficients: (1 not defined because of singularities)
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.197157   0.101940 -11.744 < 2e-16 ***
## q1female         0.090040   0.033919   2.655  0.00794 **
## q225-30          0.241126   0.073335   3.288  0.00101 **
## q231-60          0.333047   0.072492   4.594 4.34e-06 ***
## q260 and above   0.540629   0.079775   6.777 1.23e-11 ***
## q3engineer and programmer 0.094992   0.036043   2.636  0.00840 **
## q3medical professional -0.106060   0.052852  -2.007  0.04478 *
## q3business man/woman  -0.033418   0.064195  -0.521  0.60267
## q3student         0.124732   0.075541   1.651  0.09870 .
## q4khartoum north    0.066454   0.030845   2.154  0.03121 *
```

```

## q4om durman -0.036385 0.030543 -1.191 0.23355
## q53-5 0.172905 0.042750 4.045 5.24e-05 ***
## q56-10 0.143069 0.045779 3.125 0.00178 **
## q510 and above 0.062685 0.044455 1.410 0.15852
## q6Sudan 0.006016 0.049855 0.121 0.90395
## q6Japan -0.131086 0.093352 -1.404 0.16026
## q6Germany -0.121588 0.174106 -0.698 0.48496
## q6Czech 0.263176 0.170955 1.539 0.12370
## q7Toyota 0.093107 0.116587 0.799 0.42452
## q7Giad NA NA NA NA
## q7Kia 0.016951 0.054263 0.312 0.75475
## q7Skoda -0.329004 0.234369 -1.404 0.16038
## q7Mitsubishi -0.114383 0.163030 -0.702 0.48293
## q7Mercedes -0.029427 0.134577 -0.219 0.82691
## q8Tuson -0.036117 0.105467 -0.342 0.73202
## q8Visto -0.146839 0.086363 -1.700 0.08908 .
## q8Click -0.085358 0.051891 -1.645 0.09998 .
## q8Fabia 0.235755 0.147415 1.599 0.10976
## q8Lancer 0.200626 0.163176 1.230 0.21888
## q8Corolla 0.117864 0.083820 1.406 0.15968
## q8Mercedes -0.135766 0.145252 -0.935 0.34995
## q8Hilux 0.049825 0.127754 0.390 0.69653
## q8Land cruiser 0.014788 0.128780 0.115 0.90858
## q96-10 -0.058141 0.031668 -1.836 0.06636 .
## q911-20 -0.262621 0.036207 -7.253 4.07e-13 ***
## q921 and above -0.683216 0.099182 -6.888 5.64e-12 ***
## q10Medium 0.110389 0.068099 1.621 0.10502
## q10Large 0.018442 0.123722 0.149 0.88150
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 15672 on 13887 degrees of freedom
## Residual deviance: 15356 on 13851 degrees of freedom
## (797 observations deleted due to missingness)
## AIC: 25460
##
## Number of Fisher Scoring iterations: 6
## Analysis of Deviance Table
##
## Model: poisson, link: log

```

```
##
## Response: q11
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                13887      15672
## q1      1      5.575      13886      15667 0.0182203 *
## q2      3     75.429      13883      15592 2.932e-16 ***
## q3      4     19.951      13879      15572 0.0005106 ***
## q4      2      8.870      13877      15563 0.0118532 *
## q5      3     22.810      13874      15540 4.423e-05 ***
## q6      4     16.264      13870      15524 0.0026846 **
## q7      5      7.548      13865      15516 0.1829537
## q8      9     56.384      13856      15460 6.628e-09 ***
## q9      3    100.381      13853      15359 < 2.2e-16 ***
## q10     2      3.471      13851      15356 0.1763213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## glm.nb(formula = q11 ~ q1 + q2 + q3 + q4 + q5 + q6 + q7 + q8 +
##      q9 + q10, data = df, init.theta = 1.491152932, link = log)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -1.2349  -0.9139  -0.8250   0.5037   3.9798
##
## Coefficients: (1 not defined because of singularities)
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.19918    0.11461 -10.463  < 2e-16 ***
## q1female         0.09546    0.03915   2.439 0.014745 *
## q225-30          0.23337    0.08189   2.850 0.004377 **
## q231-60          0.32805    0.08076   4.062 4.86e-05 ***
## q260 and above   0.54006    0.08964   6.025 1.69e-09 ***
## q3engineer and programmer 0.09302    0.04179   2.226 0.026003 *
## q3medical professional -0.11008    0.06016  -1.830 0.067286 .
## q3business man/woman  -0.03383    0.07360  -0.460 0.645771
## q3student         0.10605    0.08601   1.233 0.217576
## q4khartoum north    0.06813    0.03556   1.916 0.055356 .
## q4om durman       -0.03453    0.03493  -0.989 0.322882
```

```

## q53-5          0.17259    0.04857    3.553 0.000381 ***
## q56-10         0.14070    0.05213    2.699 0.006947 **
## q510 and above 0.05890    0.05050    1.166 0.243523
## q6Sudan        0.01212    0.05753    0.211 0.833121
## q6Japan        -0.13506    0.10675   -1.265 0.205834
## q6Germany      -0.12199    0.19848   -0.615 0.538804
## q6Czech        0.29051    0.20213    1.437 0.150644
## q7Toyota       0.09961    0.13379    0.745 0.456551
## q7Giad          NA          NA          NA      NA
## q7Kia          0.01754    0.06221    0.282 0.778041
## q7Skoda        -0.35624    0.27382   -1.301 0.193268
## q7Mitsubishi   -0.11212    0.18462   -0.607 0.543644
## q7Mercedes     -0.03358    0.15426   -0.218 0.827692
## q8Tuson        -0.03518    0.12087   -0.291 0.770990
## q8Visto        -0.14062    0.09755   -1.442 0.149408
## q8Click        -0.07634    0.05905   -1.293 0.196065
## q8Fabia        0.24298    0.16882    1.439 0.150070
## q8Lancer       0.20584    0.18477    1.114 0.265281
## q8Corolla      0.12023    0.09641    1.247 0.212395
## q8Mercedes     -0.11173    0.16277   -0.686 0.492454
## q8Hilux        0.04320    0.14706    0.294 0.768958
## q8Land cruiser 0.01397    0.14785    0.094 0.924714
## q96-10         -0.05959    0.03661   -1.628 0.103568
## q911-20        -0.26560    0.04137   -6.420 1.37e-10 ***
## q921 and above -0.67875    0.10807   -6.280 3.38e-10 ***
## q10Medium      0.11610    0.07716    1.505 0.132438
## q10Large       0.02540    0.14130    0.180 0.857328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.4912) family taken to be 1)
##
##      Null deviance: 12025  on 13900  degrees of freedom
## Residual deviance: 11780  on 13864  degrees of freedom
## (784 observations deleted due to missingness)
## AIC: 25003
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  1.4912
##              Std. Err.: 0.0930

```

```
##
## 2 x log-likelihood: -24927.0240

## Warning in anova.negbin(model1, test = "Chisq"): tests made without re-estimating
'theta'

## Analysis of Deviance Table
##
## Model: Negative Binomial(1.4912), link: log
##
## Response: q11
##
## Terms added sequentially (first to last)
##
##
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
## NULL			13900	12025	
## q1	1	4.350	13899	12021	0.037017 *
## q2	3	57.985	13896	11963	1.583e-12 ***
## q3	4	15.236	13892	11948	0.004236 **
## q4	2	6.479	13890	11941	0.039180 *
## q5	3	17.668	13887	11924	0.000515 ***
## q6	4	12.916	13883	11911	0.011695 *
## q7	5	6.177	13878	11904	0.289414
## q8	9	43.476	13869	11861	1.763e-06 ***
## q9	3	77.650	13866	11783	< 2.2e-16 ***
## q10	2	2.868	13864	11780	0.238362

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## summary.model1..coef.summary.model1..coef...4.....0.1..4.
## (Intercept) 1.274549e-25
## q1female 1.474478e-02
## q225-30 4.376668e-03
## q231-60 4.861102e-05
## q260 and above 1.691923e-09
## q3engineer and programmer 2.600313e-02
## q3medical professional 6.728634e-02
## q4khartoum north 5.535561e-02
## q53-5 3.807002e-04
## q56-10 6.947236e-03
## q911-20 1.366571e-10
## q921 and above 3.375561e-10
##
```

```
## Call:
## glm(formula = q11 ~ q1 + q2 + q3 + q4 + q5 + q6 + q7 + q8 + q9 +
##      q10, family = "quasipoisson", data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -1.4031   -0.9828   -0.8755    0.6040    5.3766
##
## Coefficients: (1 not defined because of singularities)
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.19557     0.11760  -10.166  < 2e-16 ***
## q1female         0.09119     0.03914   2.330  0.019808 *
## q225-30          0.23977     0.08460   2.834  0.004602 **
## q231-60          0.33298     0.08363   3.982  6.88e-05 ***
## q260 and above   0.54005     0.09203   5.868  4.51e-09 ***
## q3engineer and programmer 0.09464     0.04159   2.276  0.022882 *
## q3medical professional -0.10906     0.06098  -1.789  0.073691 .
## q3business man/woman  -0.03301     0.07407  -0.446  0.655811
## q3student         0.12240     0.08715   1.405  0.160173
## q4khartoum north    0.06489     0.03558   1.824  0.068201 .
## q4om durman        -0.03660     0.03524  -1.039  0.298980
## q53-5             0.17168     0.04931   3.482  0.000499 ***
## q56-10            0.14023     0.05280   2.656  0.007926 **
## q510 and above     0.06110     0.05127   1.192  0.233419
## q6Sudan           0.00615     0.05752   0.107  0.914855
## q6Japan           -0.12915     0.10771  -1.199  0.230541
## q6Germany         -0.12846     0.20085  -0.640  0.522470
## q6Czech           0.26304     0.19725   1.334  0.182374
## q7Toyota          0.09639     0.13445   0.717  0.473432
## q7Giad            NA          NA      NA      NA
## q7Kia             0.01664     0.06260   0.266  0.790460
## q7Skoda           -0.33095     0.27053  -1.223  0.221230
## q7Mitsubishi      -0.11802     0.18807  -0.628  0.530332
## q7Mercedes        -0.02907     0.15525  -0.187  0.851457
## q8Tuson           -0.03350     0.12175  -0.275  0.783202
## q8Visto           -0.14534     0.09964  -1.459  0.144709
## q8Click           -0.08378     0.05987  -1.399  0.161742
## q8Fabia           0.23980     0.17020   1.409  0.158889
## q8Lancer          0.20418     0.18824   1.085  0.278072
## q8Corolla         0.11357     0.09662   1.175  0.239837
## q8Mercedes       -0.11935     0.16661  -0.716  0.473784
## q8Hilux           0.04874     0.14747   0.331  0.741015
```

```

## q8Land cruiser          0.01474    0.14862    0.099 0.921007
## q96-10                  -0.05943    0.03653   -1.627 0.103846
## q911-20                 -0.26334    0.04178   -6.304 3.00e-10 ***
## q921 and above          -0.67992    0.11401   -5.964 2.53e-09 ***
## q10Medium                0.11030    0.07858    1.404 0.160424
## q10Large                 0.01532    0.14280    0.107 0.914596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.331282)
##
##      Null deviance: 15684  on 13900  degrees of freedom
## Residual deviance: 15369  on 13864  degrees of freedom
## (784 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
## Analysis of Deviance Table
##
## Model: quasipoisson, link: log
##
## Response: q11
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                    13900      15684
## q1      1      5.707      13899      15678 0.0384131 *
## q2      3     75.286      13896      15603 3.203e-12 ***
## q3      4     19.976      13892      15583 0.0046914 **
## q4      2      8.594      13890      15574 0.0396513 *
## q5      3     22.524      13887      15552 0.0007342 ***
## q6      4     16.360      13883      15536 0.0153289 *
## q7      5      7.622      13878      15528 0.3338633
## q8      9     55.541      13869      15472 3.698e-06 ***
## q9      3    100.125      13866      15372 3.267e-16 ***
## q10     2      3.520      13864      15369 0.2666134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```


Combine results of the three Distributions

##	=====		
##	Dependent variable:		
##	-----		
##	q11		
##	Poisson	negative	glm: quasipoisson
##		binomial	link = log
##	(1)	(2)	(3)
##	-----		
## q1female	0.09**	0.10*	0.09*
##	(0.03)	(0.04)	(0.04)
## q225-30	0.24**	0.23**	0.24**
##	(0.07)	(0.08)	(0.08)
## q231-60	0.33***	0.33***	0.33***
##	(0.07)	(0.08)	(0.08)
## q260 and above	0.54***	0.54***	0.54***
##	(0.08)	(0.09)	(0.09)
## q3engineer and programmer	0.09**	0.09*	0.09*
##	(0.04)	(0.04)	(0.04)
## q3medical professional	-0.11*	-0.11	-0.11
##	(0.05)	(0.06)	(0.06)
## q3business man/woman	-0.03	-0.03	-0.03
##	(0.06)	(0.07)	(0.07)
## q3student	0.12	0.11	0.12
##	(0.08)	(0.09)	(0.09)
## q4khartoum north	0.07*	0.07	0.06
##	(0.03)	(0.04)	(0.04)
## q4om durman	-0.04	-0.03	-0.04
##	(0.03)	(0.03)	(0.04)
## q53-5	0.17***	0.17***	0.17***
##	(0.04)	(0.05)	(0.05)
## q56-10	0.14**	0.14**	0.14**
##	(0.05)	(0.05)	(0.05)
## q510 and above	0.06	0.06	0.06
##	(0.04)	(0.05)	(0.05)
## q6Sudan	0.01	0.01	0.01
##	(0.05)	(0.06)	(0.06)
## q6Japan	-0.13	-0.14	-0.13
##	(0.09)	(0.11)	(0.11)
## q6Germany	-0.12	-0.12	-0.13

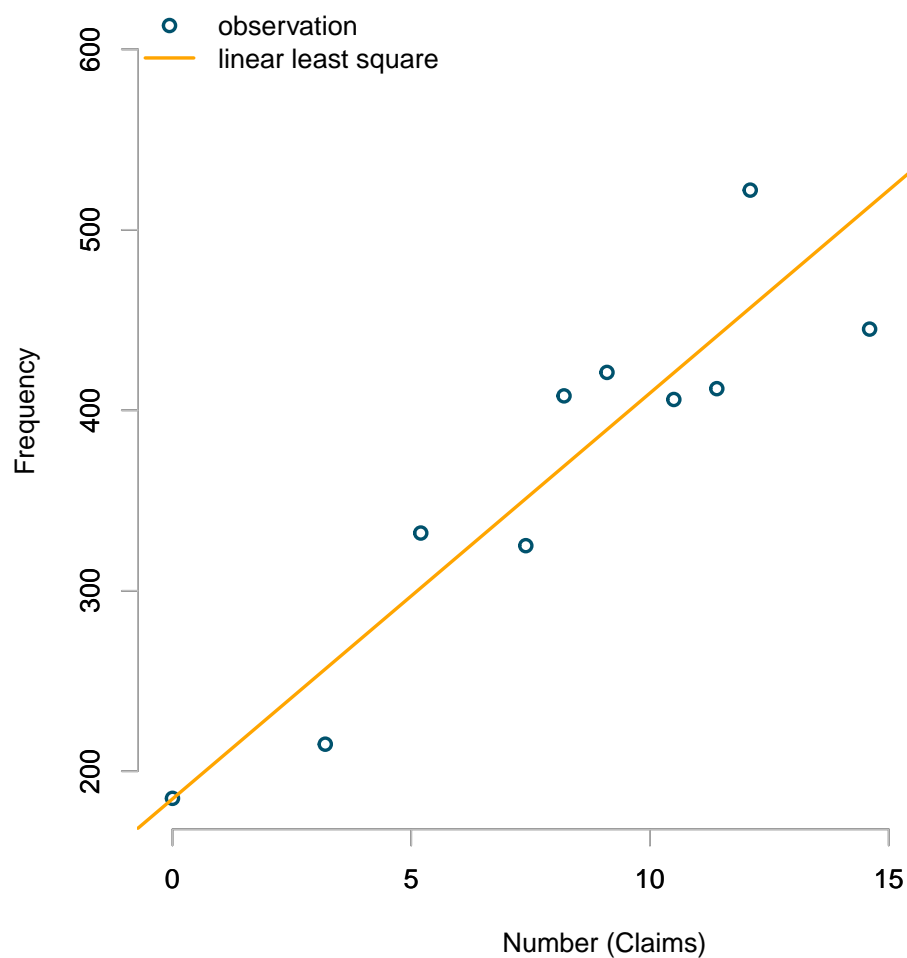
##	(0.17)	(0.20)	(0.20)
## q6Czech	0.26	0.29	0.26
##	(0.17)	(0.20)	(0.20)
## q7Toyota	0.09	0.10	0.10
##	(0.12)	(0.13)	(0.13)
## q7Giad			
##			
## q7Kia	0.02	0.02	0.02
##	(0.05)	(0.06)	(0.06)
## q7Skoda	-0.33	-0.36	-0.33
##	(0.23)	(0.27)	(0.27)
## q7Mitsubishi	-0.11	-0.11	-0.12
##	(0.16)	(0.18)	(0.19)
## q7Merceds	-0.03	-0.03	-0.03
##	(0.13)	(0.15)	(0.16)
## q8Tuson	-0.04	-0.04	-0.03
##	(0.11)	(0.12)	(0.12)
## q8Visto	-0.15	-0.14	-0.15
##	(0.09)	(0.10)	(0.10)
## q8Click	-0.09	-0.08	-0.08
##	(0.05)	(0.06)	(0.06)
## q8Fabia	0.24	0.24	0.24
##	(0.15)	(0.17)	(0.17)
## q8Lancer	0.20	0.21	0.20
##	(0.16)	(0.18)	(0.19)
## q8Corolla	0.12	0.12	0.11
##	(0.08)	(0.10)	(0.10)
## q8Merceds	-0.14	-0.11	-0.12
##	(0.15)	(0.16)	(0.17)
## q8Hilux	0.05	0.04	0.05
##	(0.13)	(0.15)	(0.15)
## q8Land cruiser	0.01	0.01	0.01
##	(0.13)	(0.15)	(0.15)
## q96-10	-0.06	-0.06	-0.06
##	(0.03)	(0.04)	(0.04)
## q911-20	-0.26***	-0.27***	-0.26***
##	(0.04)	(0.04)	(0.04)
## q921 and above	-0.68***	-0.68***	-0.68***
##	(0.10)	(0.11)	(0.11)
## q10Medium	0.11	0.12	0.11
##	(0.07)	(0.08)	(0.08)
## q10Large	0.02	0.03	0.02

##	(0.12)	(0.14)	(0.14)
## Constant	-1.20***	-1.20***	-1.20***
##	(0.10)	(0.11)	(0.12)
## -----			
## Observations	13,888	13,901	13,901
## Log Likelihood	-12,693.21	-12,464.51	
## theta		1.49*** (0.09)	
## Akaike Inf. Crit.	25,460.41	25,003.02	
## =====			
## Note:		*p<0.05; **p<0.01; ***p<0.001	

Plotting and comparing frequency models

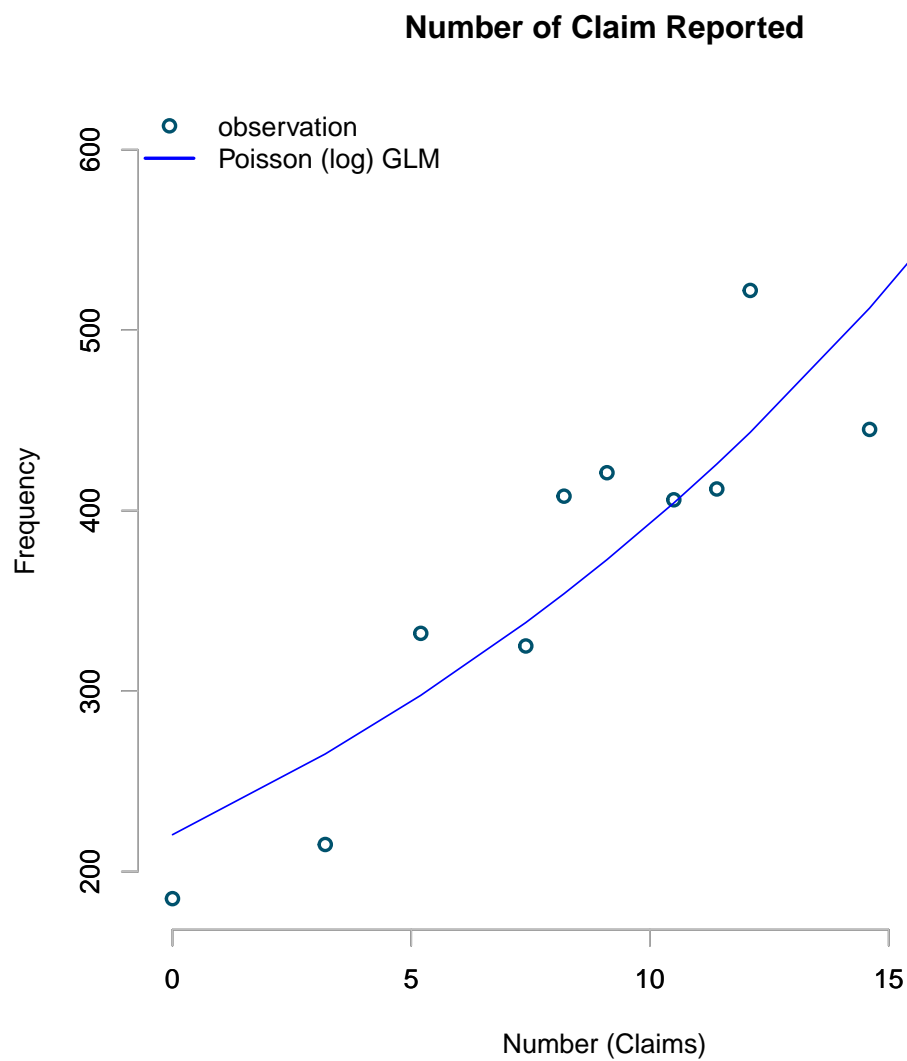
```
##  
## Call:  
## lm(formula = frequency ~ claim, data = reported)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -67.970 -26.810  -4.563  30.720  65.243   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  184.682     25.039   7.376 2.38e-05 ***  
## claim        22.486       2.287   9.832 1.86e-06 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 40.48 on 10 degrees of freedom  
## Multiple R-squared:  0.9063, Adjusted R-squared:  0.8969   
## F-statistic: 96.67 on 1 and 10 DF,  p-value: 1.856e-06
```

Number of Claim Reported



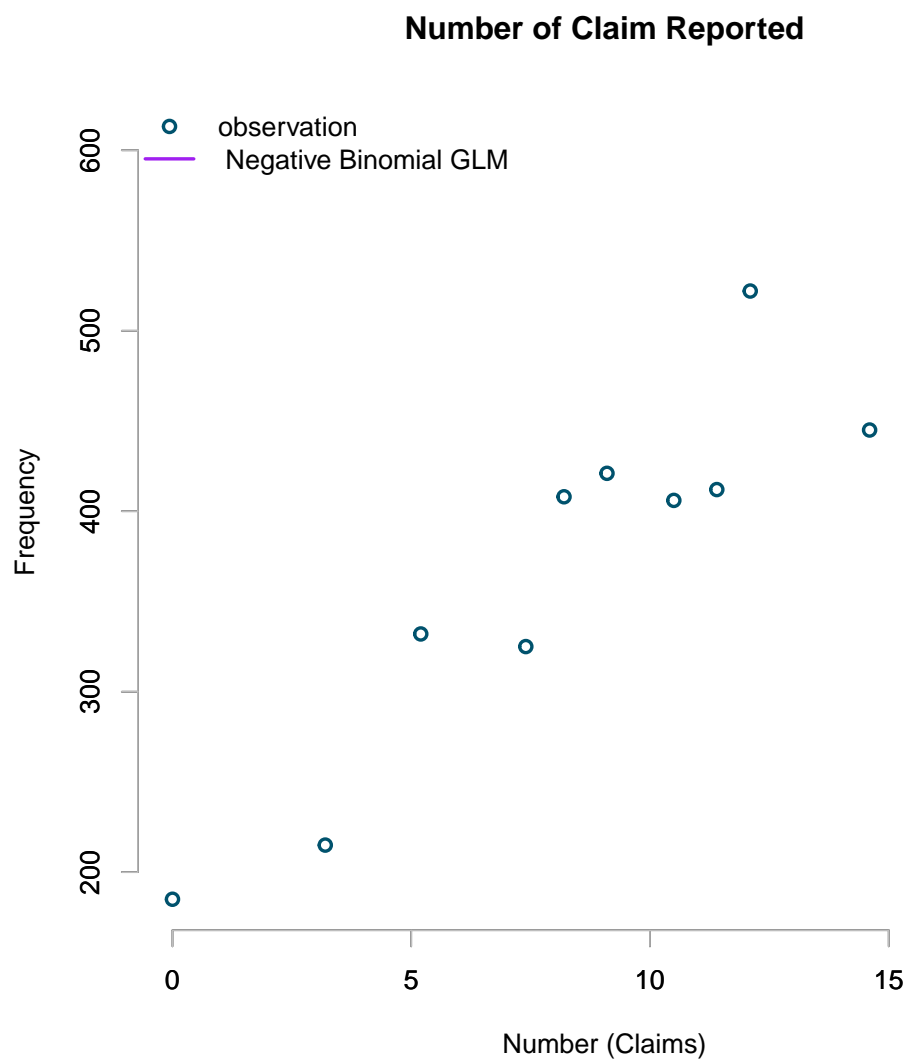
```
##
## Call:
## glm(formula = frequency ~ claim, family = poisson(link = "log"),
##      data = reported)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1870  -1.3855  -0.5965   2.0769   3.6313
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.395581   0.035643  151.38  <2e-16 ***
## claim        0.057754   0.002927   19.73  <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: 460.137  on 11  degrees of freedom
## Residual deviance:  58.557  on 10  degrees of freedom
## AIC: 155.95
##
## Number of Fisher Scoring iterations: 4
```

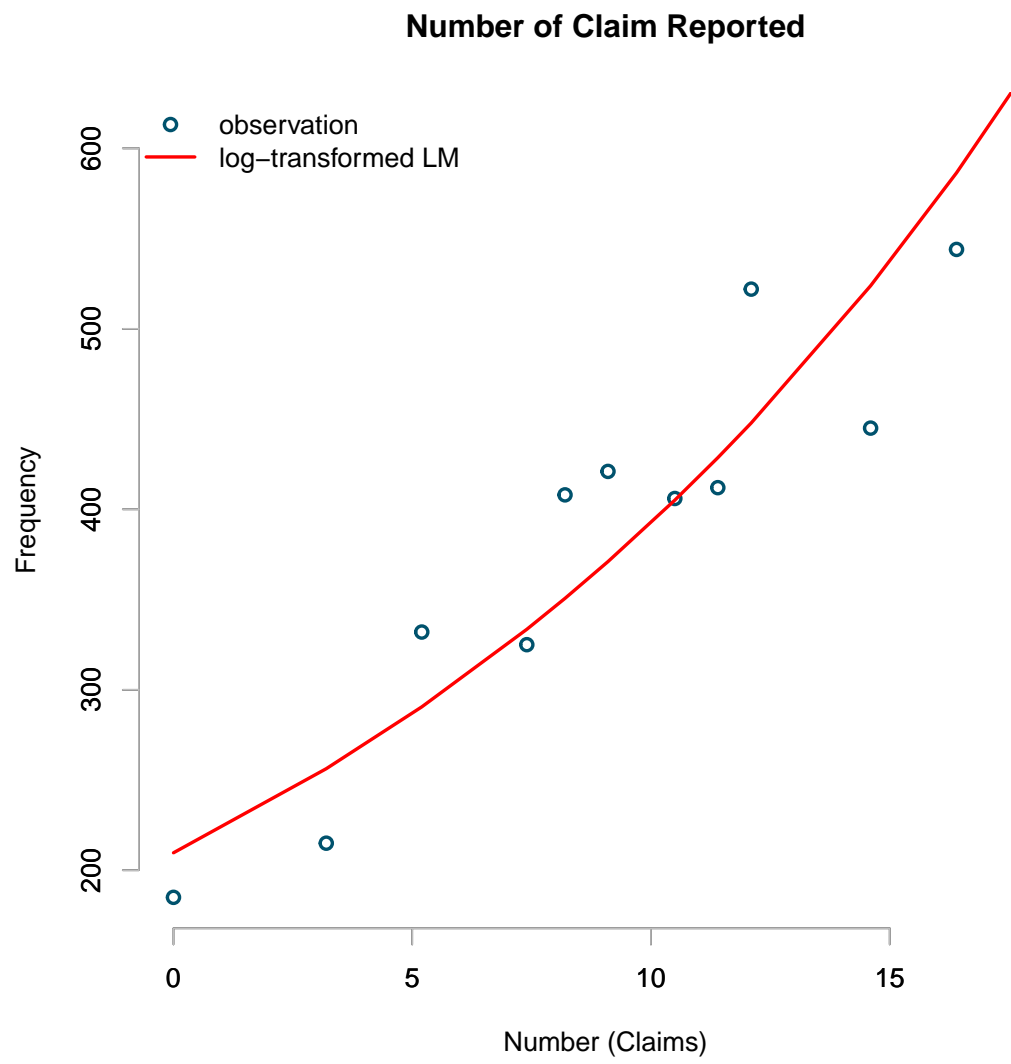


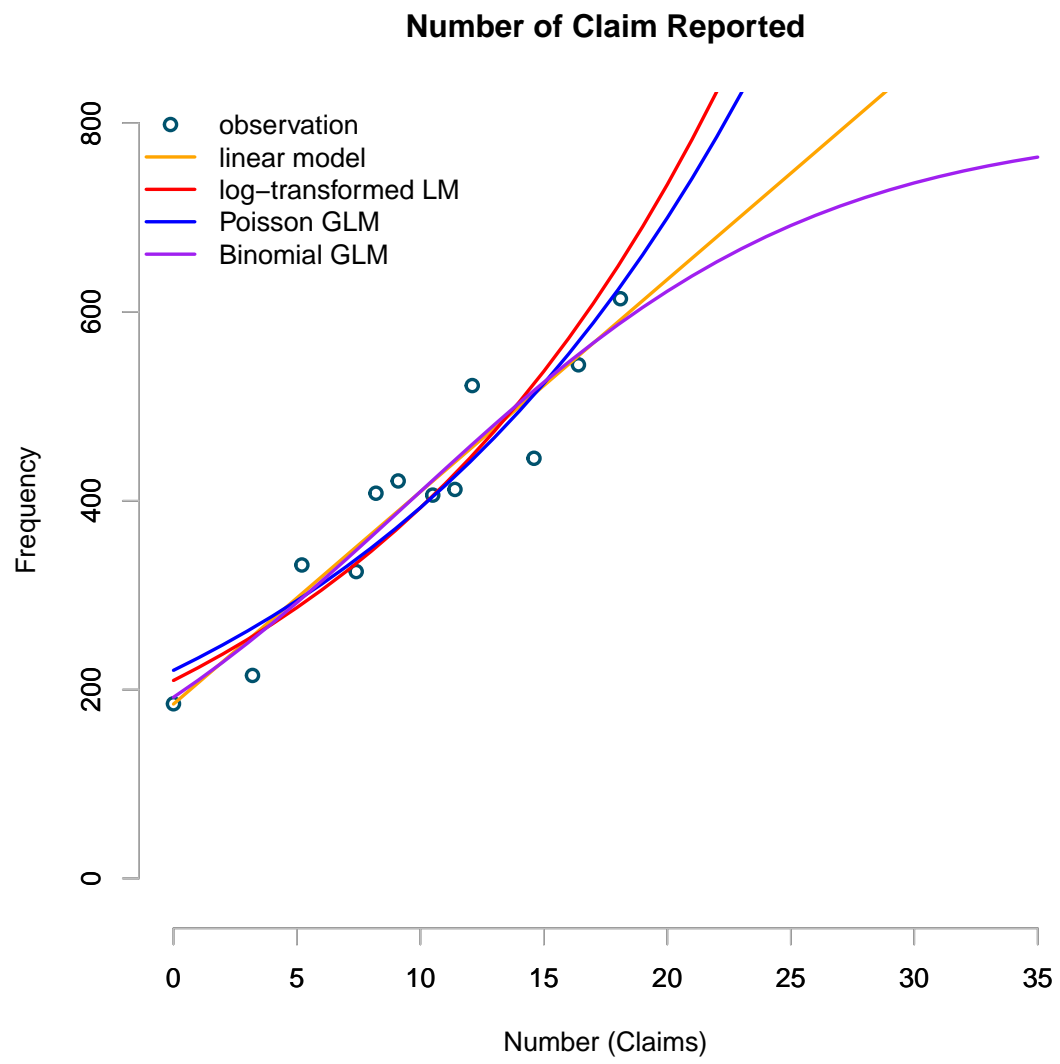
```
##
## Call:
## glm(formula = cbind(frequency, opportunity) ~ claim, family = binomial(link = "logit")
##      data = reported)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2276  -1.7614  -0.7076   2.3871   4.4963
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.153631   0.047975  -24.05   <2e-16 ***
## claim        0.120225   0.004419   27.20   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 909.399  on 11  degrees of freedom
## Residual deviance:  91.051  on 10  degrees of freedom
## AIC: 179.46
##
## Number of Fisher Scoring iterations: 3
```



```
##
## Call:
## glm(formula = log(frequency) ~ claim, family = gaussian(link = "identity"),
##      data = reported)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -0.16742  -0.07946  -0.02447   0.13633   0.16145
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.337369   0.079241  67.356 1.27e-14 ***
## claim        0.062715   0.007238   8.665 5.81e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.01640839)
##
##      Null deviance: 1.39610  on 11  degrees of freedom
## Residual deviance: 0.16408  on 10  degrees of freedom
## AIC: -11.453
##
## Number of Fisher Scoring iterations: 2
```



For Claim Severity Estimations

Perform serverity distribution for GLM using Gamma and log Normal Distributions

```
## Error in glm.fit(x = structure(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, :  
NA/NaN/Inf in 'y'  
## Error in summary(model2): object 'model2' not found  
## Error in anova(model2, test = "Chisq"): object 'model2' not found  
## Error in eval(family$initialize): non-positive values not allowed for the  
'gamma' family  
## Error in summary(model3): object 'model3' not found  
## Error in summary.glm(model2): object 'model2' not found  
## Error in is.data.frame(x): object 'freq_results_model2' not found
```

Combine results of the three Distributions

```
## Error in .stargazer.wrap(..., type = type, title = title, style = style,  
: object 'model2' not found
```