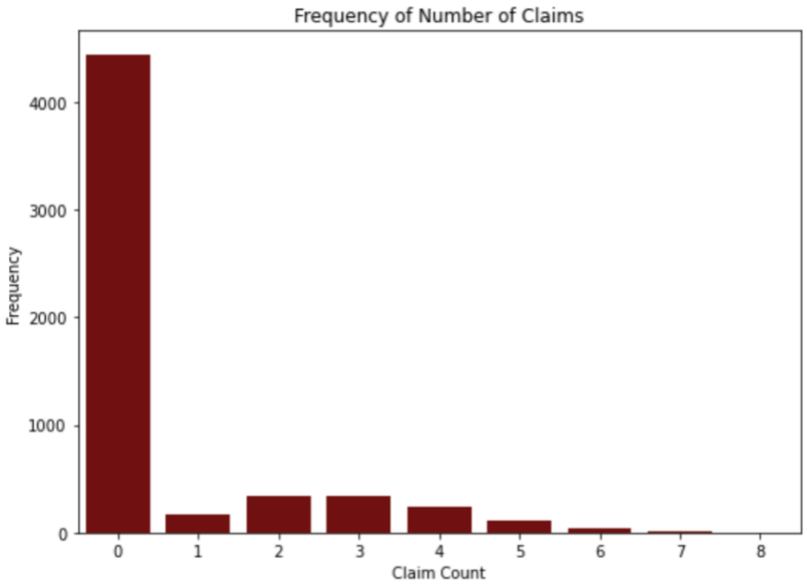
MSCA31010: Linear & Non-Linear Models Winter Quarter 2023

Assignment 2

Question 1 (20 points)

a) (10 points). Please generate a vertical bar chart to show the frequency of the number of claims.



b) (10 points). What is the log-likelihood value, the Akaike Information Criterion (AIC) value, and the Bayesian Information Criterion (BIC) value of the Intercept-only model?

```
Log-likelihood value :  -9202.190712554877
Akaike Information Criterion (AIC) value :  18406.381425109754
Bayesian Information Criterion (BIC) value :  18413.032274685982
```

Question 2 (30 points)

a) (10 points). Please provide a summary report of the Forward Selection in a table. The report should include (1) the step number, (2) the predictor entered, (3) the number of non-aliased parameters in the current model, (4) the log-likelihood value of the current model, (5) the Deviance Chi-squares statistic between the current and the previous models, (6) the corresponding Deviance Degree of Freedom, and (7) the corresponding Chi-square significance.

Step	Predictor	Non-Aliased Parameters	Log-Likelihood	Deviance Chi-Squares	Degrees of Freedom	Chi-Square Significance
0	INTERCEPT	1	-9202.190713	NaN	NaN	NaN
1	URBANICITY	2	-8796.722613	810.9361989	1	2.26E-178
2	EDUCATION	6	-8488.805338	615.8345498	4	5.80E-132
3	MVR_PTS	7	-8349.865822	277.8790324	1	2.18E-62
4	CAR_TYPE	12	-8234.98562	229.7604031	5	1.20E-47
5	TRAVTIME	13	-8163.342263	143.2867143	1	5.09E-33
6	CAR_USE	14	-8097.875925	130.9326765	1	2.56E-30
7	KIDSDRIV	15	-8049.321029	97.10979118	1	6.56E-23
8	INCOME	16	-8000.010438	98.62118359	1	3.06E-23
9	REVOKED	17	-7958.349167	83.32254054	1	6.97E-20
10	TIF	18	-7921.74615	73.20603465	1	1.17E-17
11	PARENT1	19	-7887.529214	68.43387178	1	1.31E-16
12	BLUEBOOK	20	-7868.500958	38.05651294	1	6.87E-10
13	MSTATUS	21	-7859.308943	18.38402992	1	1.81E-05
14	HOMEKIDS	22	-7854.368864	9.880158118	1	0.001670706

b) (10 points). Our final model is the model when the Forward Selection ends. What are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) of your final model?

```
Akaike Information Criterion (AIC) value :  15752.737727294354
Bayesian Information Criterion (BIC) value :  15899.05641797139
```

c) (10 points). Please show a table of the complete set of parameters of your final model (including the aliased parameters). Besides the parameter estimates, please also include the standard errors, the 95% asymptotic confidence intervals, and the exponentiated parameter estimates. Conventionally, aliased parameters have zero standard errors and confidence intervals.

	Estimate	Standard Error	Lower 95% CI	Upper 95% CI	Exponentiated
Intercept	-0.401503269	0.076406585	-0.551257424	-0.251749113	0.669313132
Highly Rural/ Rural	-1.943926116	0.081602128	-2.103863348	-1.783988883	0.143140858
Highly Urban/ Urban	0	0	0	0	1
PhD	0.139733683	0.076870341	-0.010929417	0.290396783	1.149967503
Below High Sc	0.477628984	0.055029386	0.369773369	0.585484598	1.612247203
Masters	-0.068988816	0.055440835	-0.177650856	0.039673225	0.933337119
High School	0.403390977	0.044823706	0.315538126	0.491243827	1.496892027
Bachelors	0	0	0	0	1
MVR_PTS	0.080915995	0.006531556	0.06811438	0.09371761	1.084279808
Panel Truck	-0.038398228	0.080250973	-0.195687244	0.118890788	0.962329638
Van	-0.03810268	0.066997302	-0.16941498	0.09320962	0.962614095
Sports Car	0.073539686	0.052107087	-0.028588328	0.1756677	1.07631125
Pickup	-0.234377467	0.050292752	-0.332949451	-0.135805484	0.791063159
Minivan	-0.549465362	0.050585121	-0.648610377	-0.450320347	0.577258352
SUV	0	0	0	0	1
TRAVTIME	0.011963964	0.001008998	0.009986364	0.013941563	1.012035818
Commercial	0.529038114	0.040710471	0.449247058	0.60882917	1.697298915
Private	0	0	0	0	1
KIDSDRIV	0.217801581	0.028880876	0.161196105	0.274407058	1.243340341
INCOME	-0.004413474	0.000528339	-0.005448999	-0.00337795	0.995596251
Yes	0.36863363	0.042196011	0.285930968	0.451336292	1.445757824
No	0	0	0	0	1
TIF	-0.037425322	0.00421063	-0.045678006	-0.029172638	0.96326635
Yes_PARENT1	0.203616715	0.072480888	0.061556786	0.345676645	1.225828222
No_PARENT1	0	0	0	0	1
BLUEBOOK	-0.016368588	0.002707656	-0.021675496	-0.011061681	0.983764649
No_MSTATUS	0.240218955	0.048192674	0.145763051	0.33467486	1.271527528
Yes_MSTATUS	0	0	0	0	1
HOMEKIDS	0.053730004	0.016893872	0.020618624	0.086841384	1.055199664

Question 3 (30 points)

a) (10 points). Calculate the Root Mean Squared Error, the Relative Error, the Pearson correlation, the Distance correlation, and the R-squared metrics for the Intercept-only model.

```
RMSE: 1.4635157608954519
Relative Error: 1.0759932377154993
Pearson Correlation: -0.19138309783283575
Distance Correlation: 0.22968068175486067
R-Squared: 0.00023764977519819723
```

b) (10 points). Calculate the Root Mean Squared Error, the Relative Error, the Pearson correlation, the Distance correlation, and the R-squared metrics for our final model in Question 2.

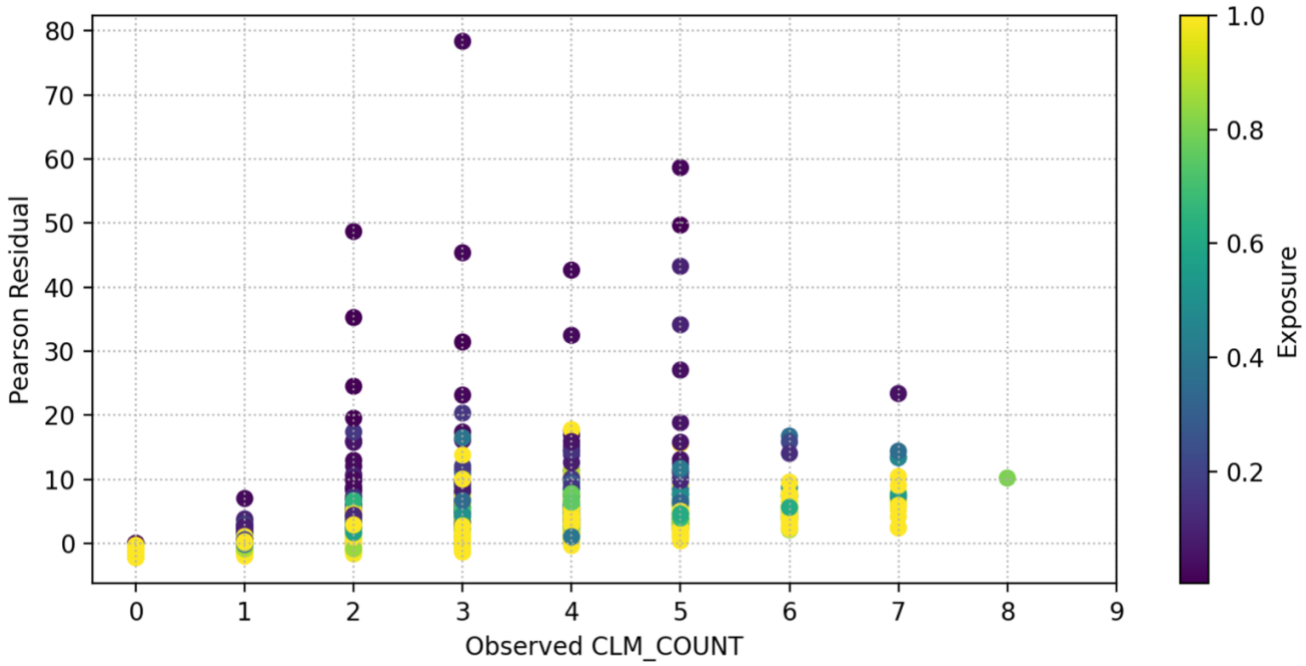
```
RMSE: 1.3946224733422352
Relative Error: 0.9770753409645692
Pearson Correlation: 0.2614873869285943
Distance Correlation: 0.2806093073756397
R-Squared: 0.0002784862880145033
```

c) (10 points) We will compare the goodness-of-fit of your model with that of the saturated model. We will calculate the Pearson Chi-Squares and the Deviance Chi-Squares statistics, their degrees of freedom, and their significance values. Based on the results, do you think your model is statistically the same as the saturated Model?

Type	Statistic	Degrees of Freedom	Significance (p-value)
Pearson	54445.80441	5693	0
Deviance	7047.060487	5693	1.74E-32

Question 4 (20 points)

a) (10 points). Plot the Pearson residuals versus the observed number of claims.



b) (10 points). Plot the Deviance residuals versus the observed number of claims.