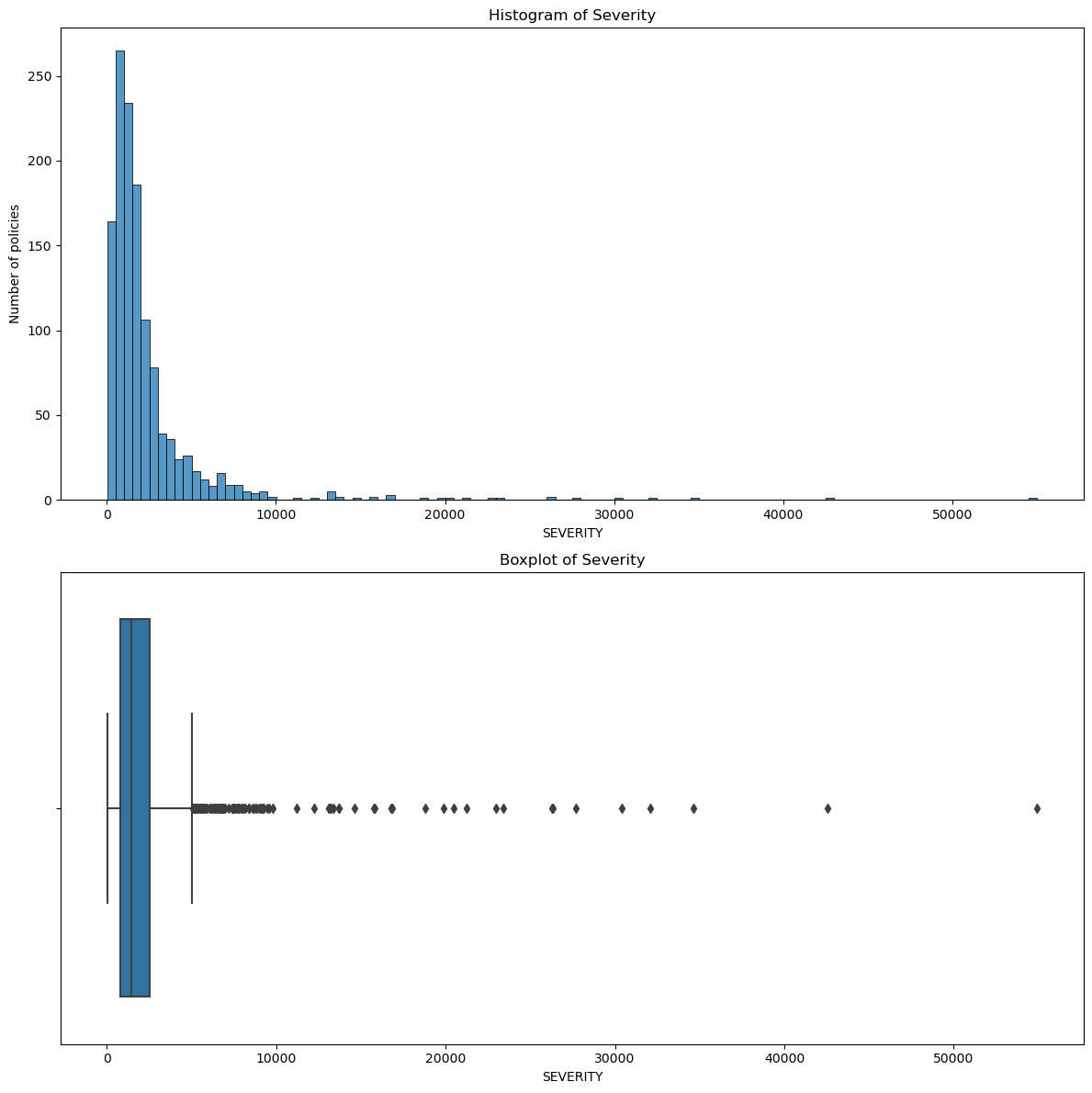
**ASSIGNMENT 4**

**QUESTION 1**

**We will first establish a baseline model for reference. To that end, we will train the Intercept-only model. This model does not include any predictors except for the Intercept term.**

1. **(10 points). Please generate a histogram and a horizontal boxplot to show the distribution of Severity. For the histogram, use a bin-width of $500 and put the number of policies on the vertical axis. Put the two graphs in the same chart where the histogram is above the boxplot.**

****

1. **(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?**

The log-likelihood value = -11171.287135771177

The Akaike Information Criterion (AIC) value = 22344.574271542355

The Bayesian Information Criterion (BIC) value = 22349.724188378488

**QUESTION 2**

**Use the Forward Selection method to build our model. The Entry Threshold is 0.01.**

1. **(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** | **# of Non-Aliased Parameters** | **Log-Likelihood** | **Deviance Chi-Square** | **Deviance Degree of Freedom** | **Chi-square significance** |
| 0 | Intercept | 1 | -11171.3 | NaN | NaN | NaN |
| 1 | + BLUEBOOK | 2 | -11157.3 | 27.90779 | 1 | 0 |
| 2 | + MSTATUS | 3 | -11145.9 | 22.85266 | 1 | 0 |
| 3 | + RED\_CAR | 4 | -11141.9 | 8.05206 | 1 | 0.00455 |
| 4 | + CAR\_TYPE | 9 | -11133.5 | 16.7553 | 5 | 0.00499 |
| 5 | + YOJ | 10 | -11129.5 | 7.96305 | 1 | 0.00477 |
| 6 | + CAR\_AGE | 11 | -11125.8 | 7.51396 | 1 | 0.00612 |

1. **(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?**

The Akaike Information Criterion (AIC) value = 22273.529459411366

The Bayesian Information Criterion (BIC) value = 22330.178544608818

1. **(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** | 7.30989 | 0.10767 | 7.09886 | 7.52092 | 1495.009 |
| **BLUEBOOK** | 0.01585 | 0.00442 | 0.00719 | 0.02451 | 1.01598 |
| **No\_MSTATUS** | 0.31905 | 0.06554 | 0.19059 | 0.4475 | 1.37581 |
| **Yes\_MSTATUS** | 0 | 0 | 0 | 0 | 1 |
| **yes\_RED\_CAR** | 0.22795 | 0.07443 | 0.08207 | 0.37382 | 1.25602 |
| **no\_RED\_CAR** | 0 | 0 | 0 | 0 | 1 |
| **Panel Truck\_CAR\_TYPE** | -0.03313 | 0.13702 | -0.30169 | 0.23543 | 0.96741 |
| **Van\_CAR\_TYPE** | -0.06256 | 0.12135 | -0.3004 | 0.17527 | 0.93935 |
| **Sports Car\_CAR\_TYPE** | 0.04995 | 0.08933 | -0.12513 | 0.22503 | 1.05122 |
| **Minivan\_CAR\_TYPE** | -0.33755 | 0.0961 | -0.5259 | -0.14919 | 0.71352 |
| **Pickup\_CAR\_TYPE** | -0.02268 | 0.08731 | -0.19381 | 0.14845 | 0.97758 |
| **SUV\_CAR\_TYPE** | 0 | 0 | 0 | 0 | 1 |
| **YOJ** | 0.01904 | 0.00728 | 0.00478 | 0.0333 | 1.01922 |
| **CAR\_AGE** | -0.01306 | 0.00522 | -0.0233 | -0.00283 | 0.98702 |

**QUESTION 3**

**We will use accuracy metrics to assess the Intercept-only model and our final model in Question 2. These metrics inform us from various perspectives how well the predicted Severity agrees with the observed Severity.**

1. **(10 points). Calculate the Root Mean Squared Error, the Relative Error, the Pearson correlation, the Distance correlation, and the Mean Absolute Proportion Error for the Intercept-only model.**

The Root Mean Squared Error = 3667.071626635712

The Relative Error = 0.9999999999999999

The Pearson correlation = -2.4563378930425157e-16

The Distance Correlation = nan

The Mean Absolute Proportion Error = 1.8861830475373358

1. **(10 points). Calculate the Root Mean Squared Error, the Relative Error, the Pearson correlation, the Distance correlation, and the Mean Absolute Proportion Error for our final model in Question 2.**

The Root Mean Squared Error = 3613.4558532056444

The Relative Error = 0.9709720320449463

The Pearson correlation = 0.170624795034319

The Distance Correlation = 0.15112268051221256

The Mean Absolute Proportion Error = 1.8216669193905304

1. **(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?**

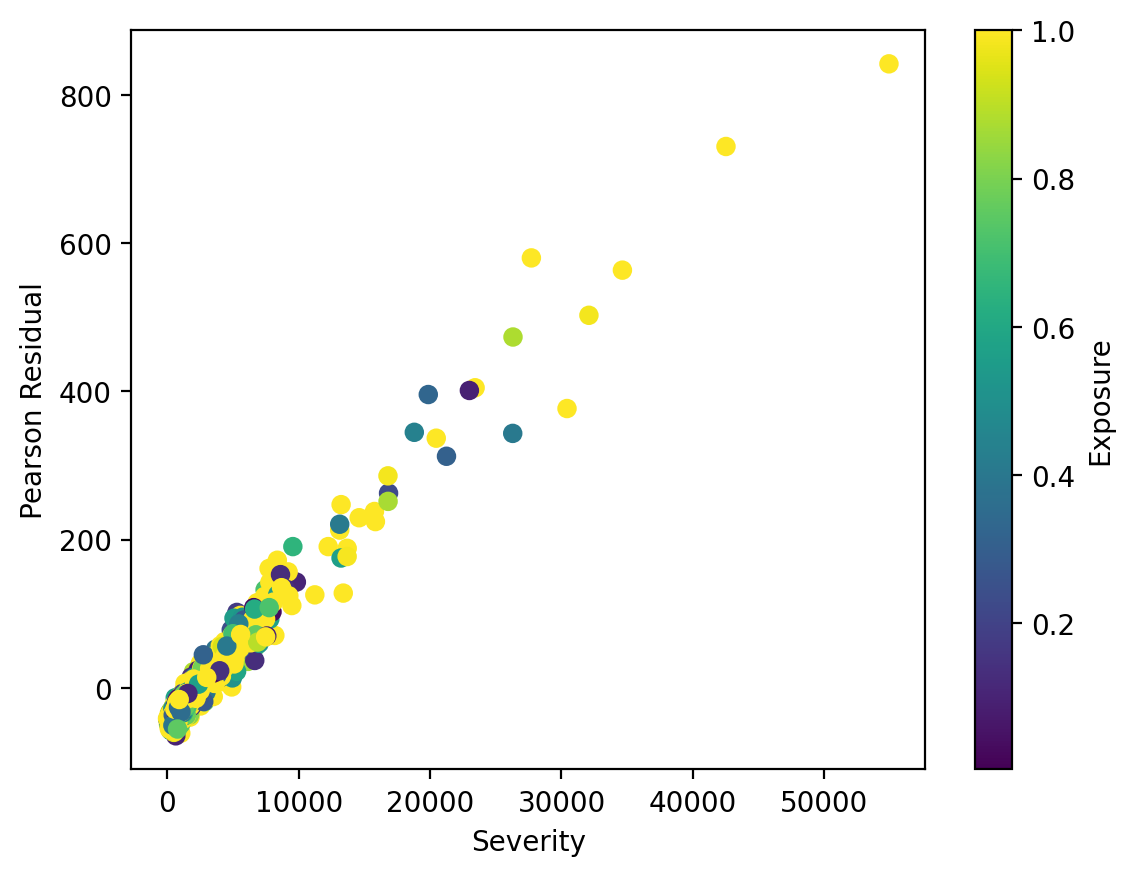
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Type** | **Statistic** | **Degrees of Freedom** | **Significance** |
| **0** | Pearson | 5887085 | 1263 | 0 |
| **1** | Deviance | 1183.202 | 1263 | 0.94621 |

Based on the results, the model is not statistically the same as the saturated model as neither are equal to 0.

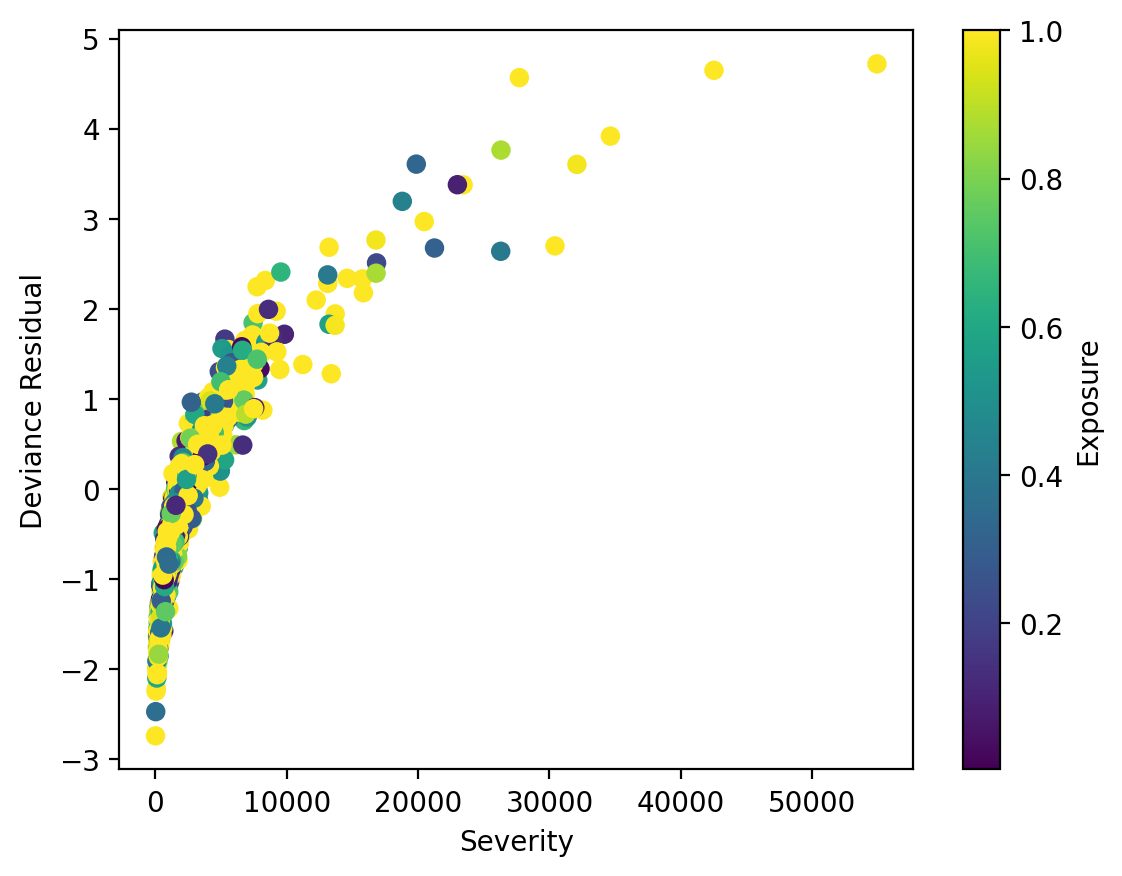
**QUESTION 4**

**You will visually assess your final model in Question 2. Please color-code the markers according to the magnitude of the Exposure value. You must properly label the axes, add grid lines, and choose appropriate tick marks to receive full credit.**

1. **(10 points). Plot the Pearson residuals versus the observed Severity.**

****

1. **(10 points). Plot the Deviance residuals versus the observed Severity.**

****