

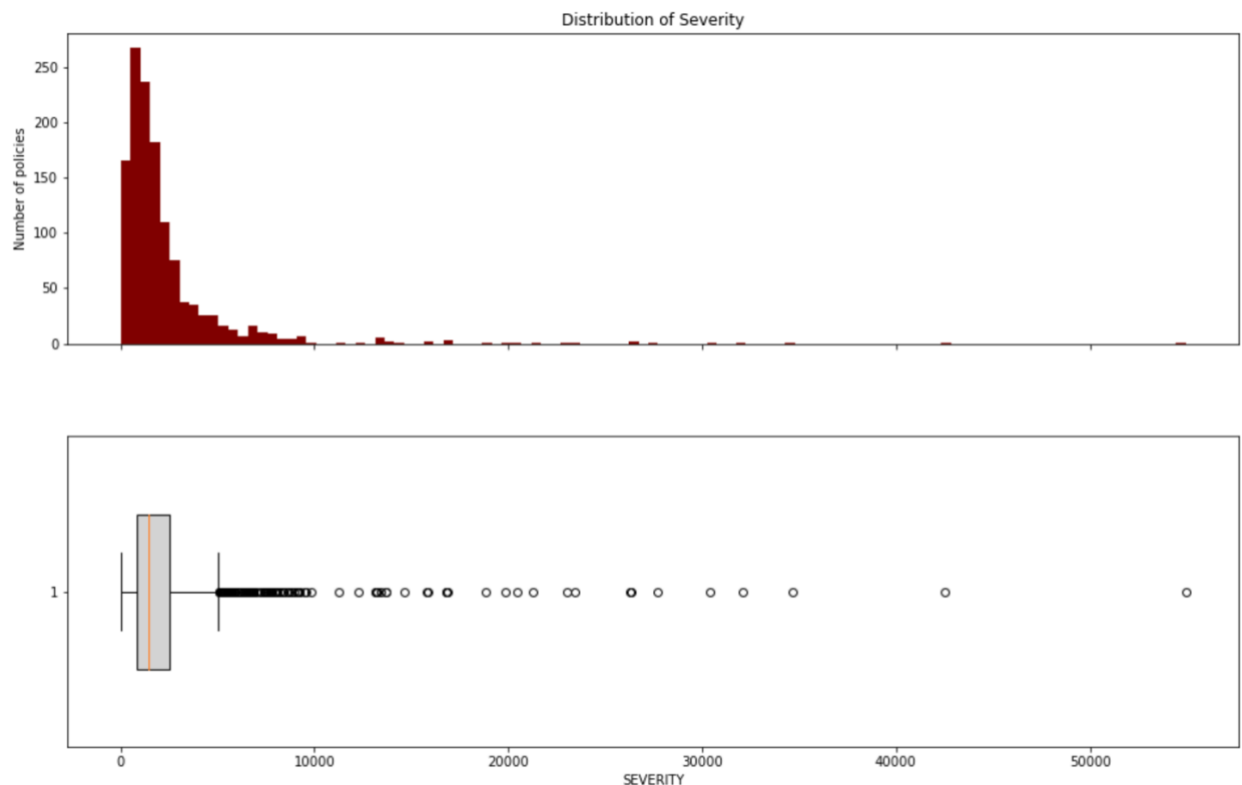
# MSCA31010: Linear & Non-Linear Models Winter Quarter 2023

## Assignment 4

---

### Question 1 (20 points)

- a) (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.



- 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 : -11171.287135771177**

**Akaike Information Criterion (AIC) value : 22344.574271542355**

**Bayesian Information Criterion (BIC) value : 22349.724188378488**

### 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	-11171.28714	NaN	NaN	NaN
1	BLUEBOOK	2	-11157.33324	27.90779111	1	1.27236E-07
2	MSTATUS	3	-11145.90691	22.8526561	1	1.74907E-06
3	RED_CAR	4	-11141.88088	8.052062275	1	0.004545189
4	CAR_TYPE	9	-11133.50323	16.75529552	5	0.004988046
5	YOJ	10	-11129.52171	7.963049681	1	0.004774191
6	CAR_AGE	11	-11125.76473	7.513957446	1	0.006122271

- 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 : 22273.529459411366**  
**Bayesian Information Criterion (BIC) value : 22330.178544608818**

- 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
<b>Intercept</b>	7.309887525	0.107671306	7.098855642	7.520919407	1495.009028
<b>BLUEBOOK</b>	0.01585228	0.004418173	0.007192821	0.024511739	1.015978594
<b>No</b>	0.3190459	0.06553979	0.190590272	0.447501529	1.375814474
<b>Yes</b>	0	0	0	0	1
<b>Yes_RED_CAR</b>	0.227948742	0.074427251	0.082074009	0.373823474	1.256020942
<b>No_RED_CAR</b>	0	0	0	0	1
<b>Panel Truck</b>	-0.033128424	0.13702136	-0.301685355	0.235428506	0.967414312
<b>Van</b>	-0.062563648	0.121346421	-0.300398264	0.175270968	0.939353273
<b>Sports Car</b>	0.049950901	0.0893275	-0.125127782	0.225029585	1.051219482
<b>Minivan</b>	-0.337545446	0.096101132	-0.525900203	-0.149190689	0.713519547
<b>Pickup</b>	-0.022679	0.087314856	-0.193812974	0.148454974	0.977576236
<b>SUV</b>	0	0	0	0	1
<b>YOJ</b>	0.019039602	0.007276402	0.004778117	0.033301088	1.019222012
<b>CAR_AGE</b>	-0.013062271	0.005222293	-0.023297777	-0.002826765	0.98702267

### 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 : 3667.071626635712**  
**Relative Error : 0.9999999999999999**  
**Pearson Correlation: -2.4563378930425157e-16**  
**Distance Correlation: None**  
**Mean Absolute Proportion Error : 1.8861830475373358**

- 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 : 3613.455853205644**  
**Relative Error : 0.9709720320449472**  
**Pearson Correlation: 0.1706247950343188**  
**Distance Correlation: 0.151122680512261**  
**Mean Absolute Proportion Error : 1.8216669193905308**

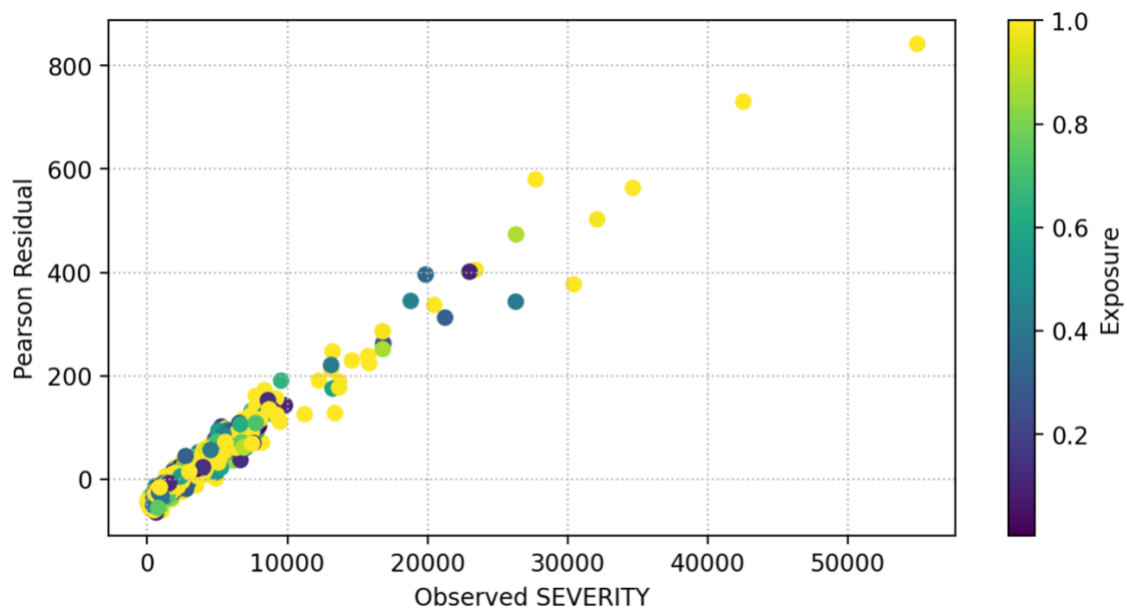
- c) (10 points) We will compare the goodness-of-fit of your model with that of the saturated model. We will calculate the Pearson Chi-Squares and the Deviance Chi-Squares statistics, their degrees of freedom, and their significance values. Based on the results, do you think your model is statistically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

Type	Statistic	Degrees of Freedom	Significance (p-value)
Pearson	5887084.650613409	1263	0
Deviance	1183.2015978998556	1263	0.946205

Our model is not statistically the same as the saturated model as neither of the statistic values are equal to zero.

#### Question 4 (20 points)

- a) (10 points). Plot the Pearson residuals versus the observed Severity.



- b) (10 points). Plot the Deviance residuals versus the observed Severity.

