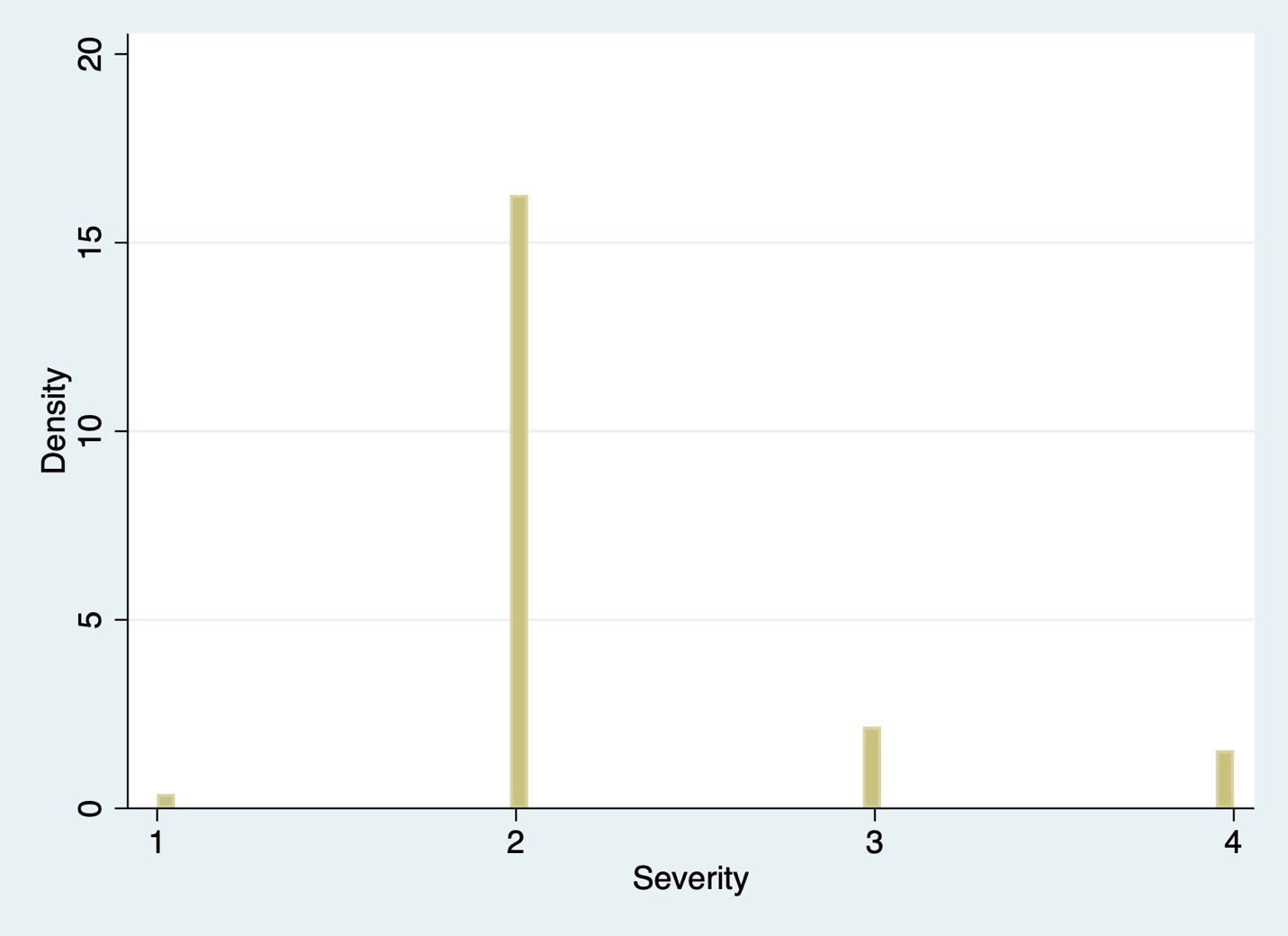
US ACCIDENTS

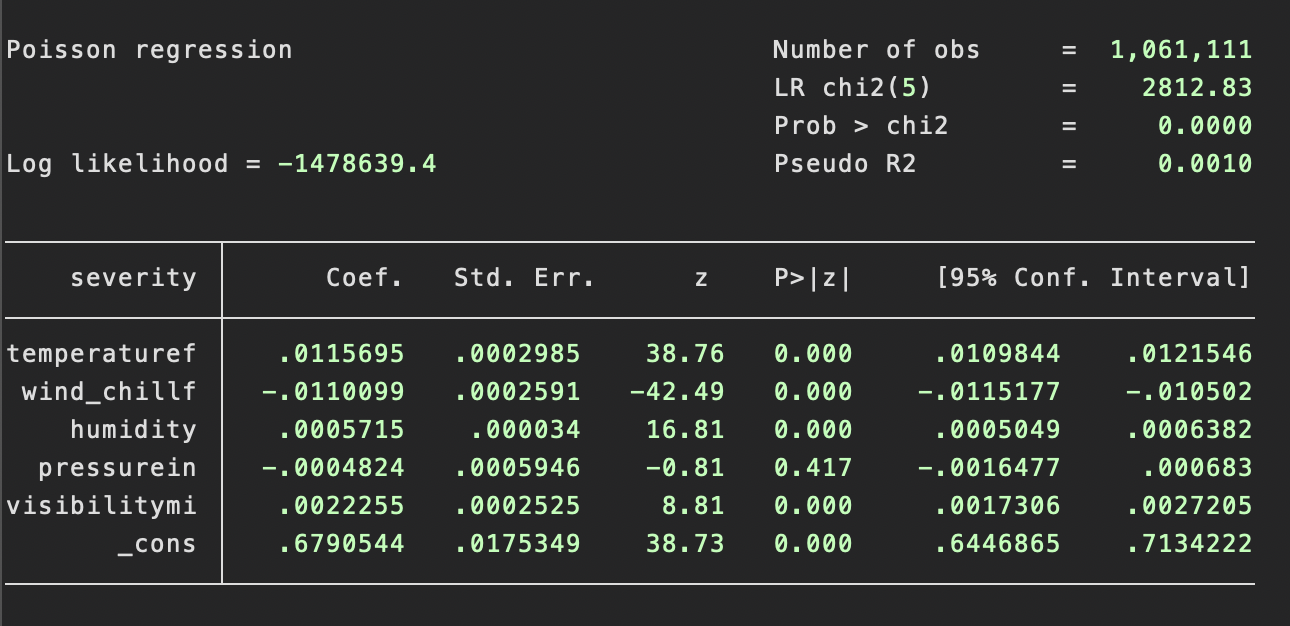
The project is to predict severity and location of US accidents. The severity of an accident is very important to health system and police. For a reported accident, police can allocate their people to the severest spot with limited number of policemen. Hospital can allocate ambulances to the most wanted accident scene. The location prediction is very important to individuals, health system, police and policy makers. Individuals can avoid road with a predicted high frequency of an accident. Police and hospitals can allocate their resources around the area with a predicted high frequency of an accident. Government can warn people in advance or maintain roads with a predicted high frequency of an accident.

I get my data from a public resource. (https://www.kaggle.com/sobhanmoosavi/us-accidents). It includes important numerical variables like temperature, wind, humidity, pressure, visibility and dummy variables like day/night, crossing/not, stop/not. I will use all the information including weather, surroundings and road information to predict the severity and location.

I make histogram of severity and find that the degree 2 is the highest.

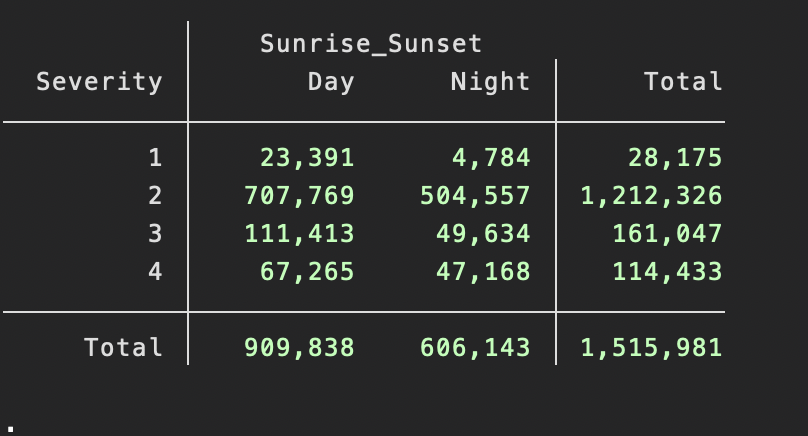


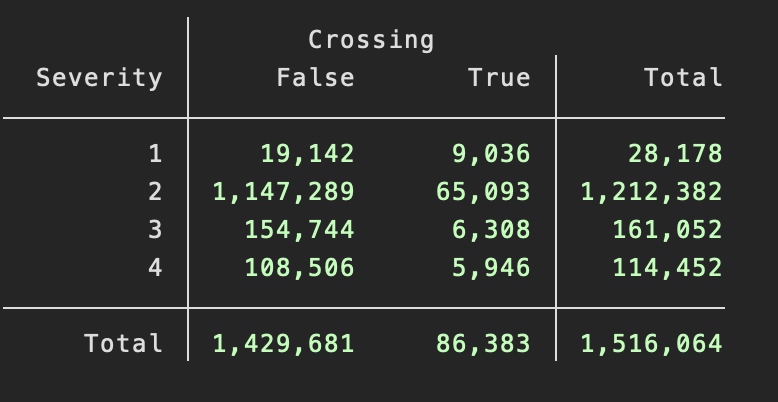
I use Linear Poisson regression to see the preliminary result of severity on other variables including temperature, wind, humidity, pressure and visibility. The result shows strong connection between those variables and severity except for pressure. We can not only predict the severity but also tell the effect of each variable on severity. E.g., if temperature goes up by 1F, there will be a 1.16% increase in the severity.



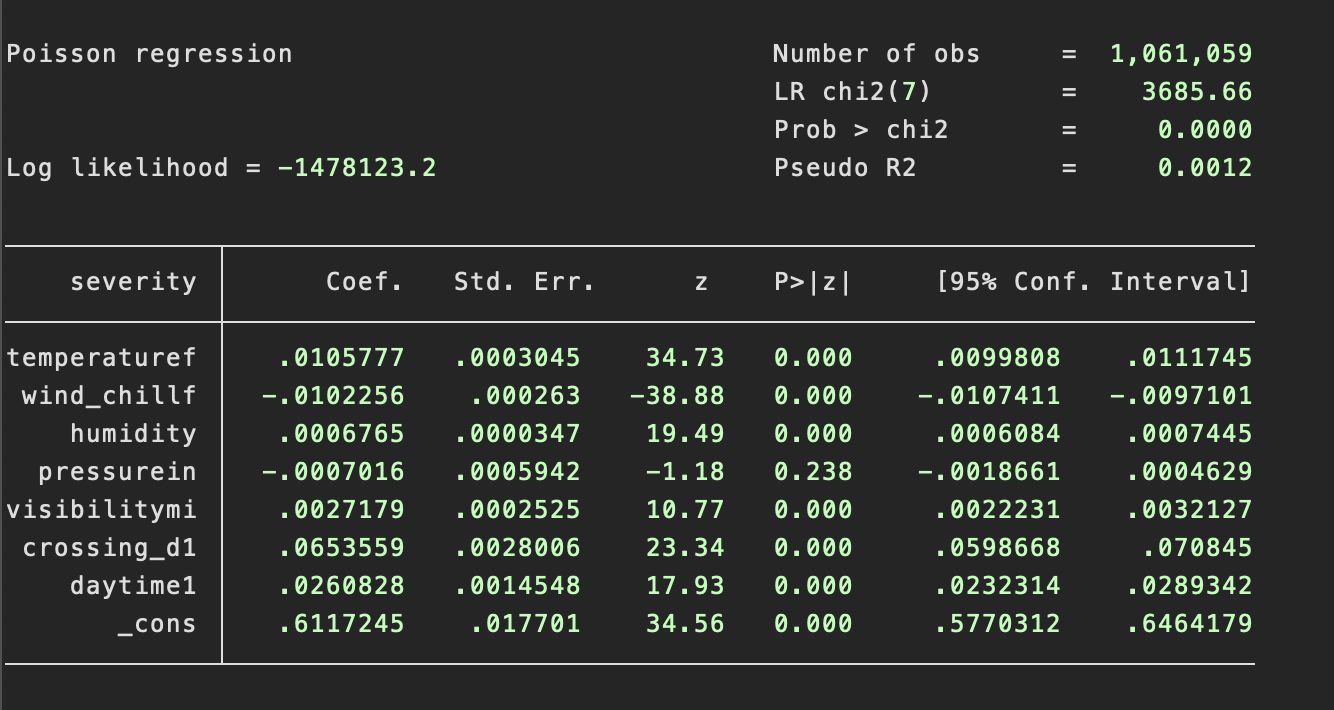
We can also include dummy variables to the analysis. I draw frequency count of severity and sunrise\_sunset. It seems accidents happened more in daytime than in nighttime. I also draw

frequency count of severity and variable describe whether there is crossing or not. Although the effect of those dummy variables are not clearly enough in this simple analysis, we can use linear Poisson regression to see whether they are relevant or not.





By adding crossing and daytime into Poisson regression, we can find the coefficients of crossing and daytime is significant and positive, indicating that those two variables are relevant to severity.



In the project, I will try various ways to analysis the data with all possible relevant regressors. Methods like non-linear Poisson regression, random forest, ensemble methods, neural network will be used.