week7

Ding Ding

October 21, 2016

Title: Analysis of factors related to fatality case on state level

Introduction

Nation's highways in motor vehicle traffice crashes result in 35,092 lives lost, 45,495 of injured survivors, and bllions of dollars in property damage[1] in 2015, which is much higher than 2014. Factors relate to fatality crashs including driver-(e.g.,age, alcohol,distracted or drowsy), road condition(e.g.,intersection, interstate,roadway departure,rural/urban),and vehicle factors(e.g.,involved truck,pedestrian/pedalcyclist,NHTSA Region,speeding, daytime/night).

Understanding the relationship between total case nubmer and the factors related can help us to characterize the risk factors for fatality cases and try to pay more attention to and avoid this risk factors. Here we performed an analysis on state level case number with to determine which factors are significantly associated with the fatality case on state level from 2010 to 2015. Using exploratory analysis and linear regression method, we show that there is a significant relationship between fatality cases level and the percentage of young/old driver in all case, the percentage of cases happend in crossection,rural place,daytime and speeding, the percentage of case involved pedalcyclist and pedestrian , and the NHTSA Region the state belongs to.

Methods

Data Collection

For our analysis , the accident data we used contains detailed, anonymized information about each of these tragic incidents of 2015, which is downloaded from NHTSA's FARS on September 9, 2016. The state/county level population data we used are estimated based on census on 2010, which is download from United States Census Bureau[2] on September 20, 2016.

Exploratory Analysis

Exploratory analysis was performed by plots and tables of data by states and year.(1) We calculated the total case number for each states from 2010 to 2015 , and we adjusted the cases with state population of the year to cases per 10,000 people.(2) For all the cases in each state per year, we calculated the percentage of factors we interested(e.g.,the percentage of cases happened in rural region in Alabama in 2010 is 60.5%). (3) We ploted and checked the distribution of the cases per 10,000 for all states in 2010 to 2015, and determine the regression method.

Statistical Modeling

To relate cases per 10,000 people to the factors related to fatality cases, we performed a standard multivariate linear regression model[3]. Model seletion was based on the results of our exploratory analysis and prior knowledge of the relationship of total cases number and the related risk factors. Coefficients were estimated with ordinary least squares and standard errors were calculated suing standard asymptotic [approximations.]4]

Reproducibility

All analyses in this manuscript are reproduced in R markdown file XXX.Rmd.

Results

The FARS data used in this analysis contains information that adjusted cases per 10,000 people for each state, percentage of single vehicle crash(CT),percentage of young driver in the cases(YOUNG),percentage of districted drivers(DIST), percentage of drowsy drivers(DOW), percentage of crashes at intersections(INTSEC), percentage of crashes at interstate(INTER), percentage of alcohol positive drivers(BAC), percentage of involving pedestrian(PED), percentage of involving pedalcyclist(PEDAL), percentage of crashes at rural region(RU), percentage of speeding(SPCRA) and percentage of crashes on daytime(TOD). We identified no missing values in the data set we collected and all measured variables were observed to be inside the standard ranges.

We first fit a regression model relating case numbers to risk factors.

Conclusions