Toan Bui–Auto Pedestrians Crashes

Source:

Question:

How has the rate of auto pedestrian incidents changed over the years? Do accidents concentrate on any time periods during the day or spread out. How can we statistically show that the rate of crashes are independent or dependent on time of day.

I chose this problem because there have been many safety improvements made on vehicles over the years and I wanted to see whether the improvements have been effective in reducing accidents. I also want to see whether or not the time of day affects the rate of auto pedestrian crashes to make an informed statistical claim on which time is most dangerous for pedestrians.

Data Description:

The data comes from recorded auto pedestrian crashes, or any accidents involving a vehicle and pedestrian, within the United States. The data is collected from 2010 to 2018.

At first, I wanted to graph rates of serious or fatal injuries during crashes to get a sense of how vehicle safety has been improved. To do this, I need to filter the types of crashes to only serious(A) and fatal(K). By doing so I get this
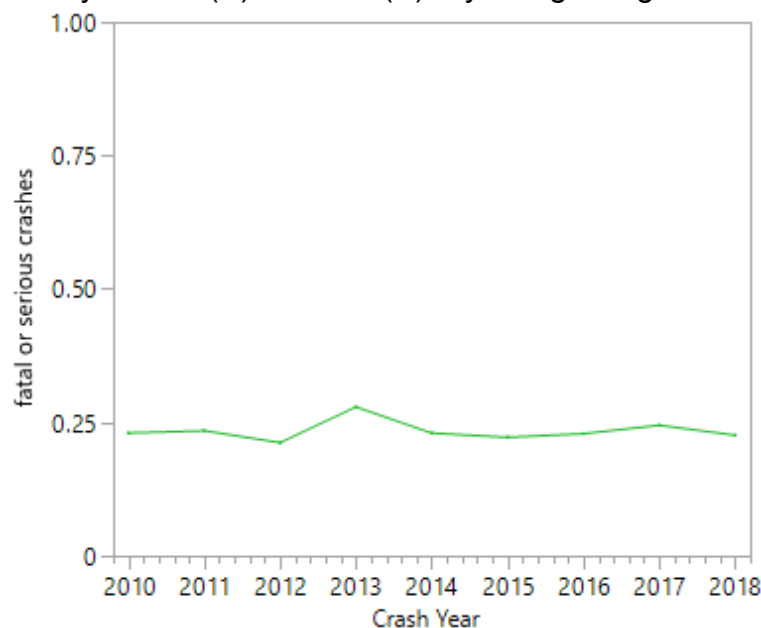
Next I also want to see the concentrations of crashes during the day. As with the previous graph, I also want to visualize the rate of serious crashes individually, but I would also like to include another graph with all crashes to test whether or not there is a difference between the two. I have set the time to 24 hour time to better organize the graph
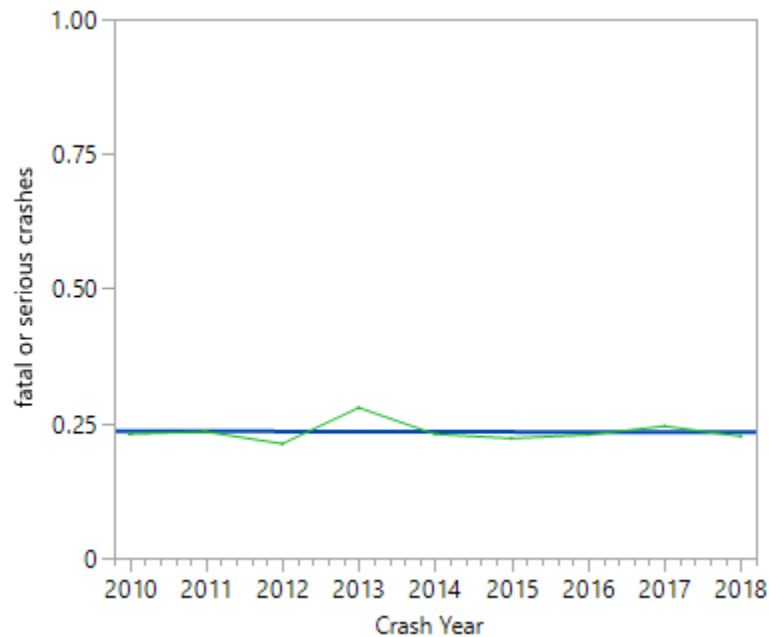

'
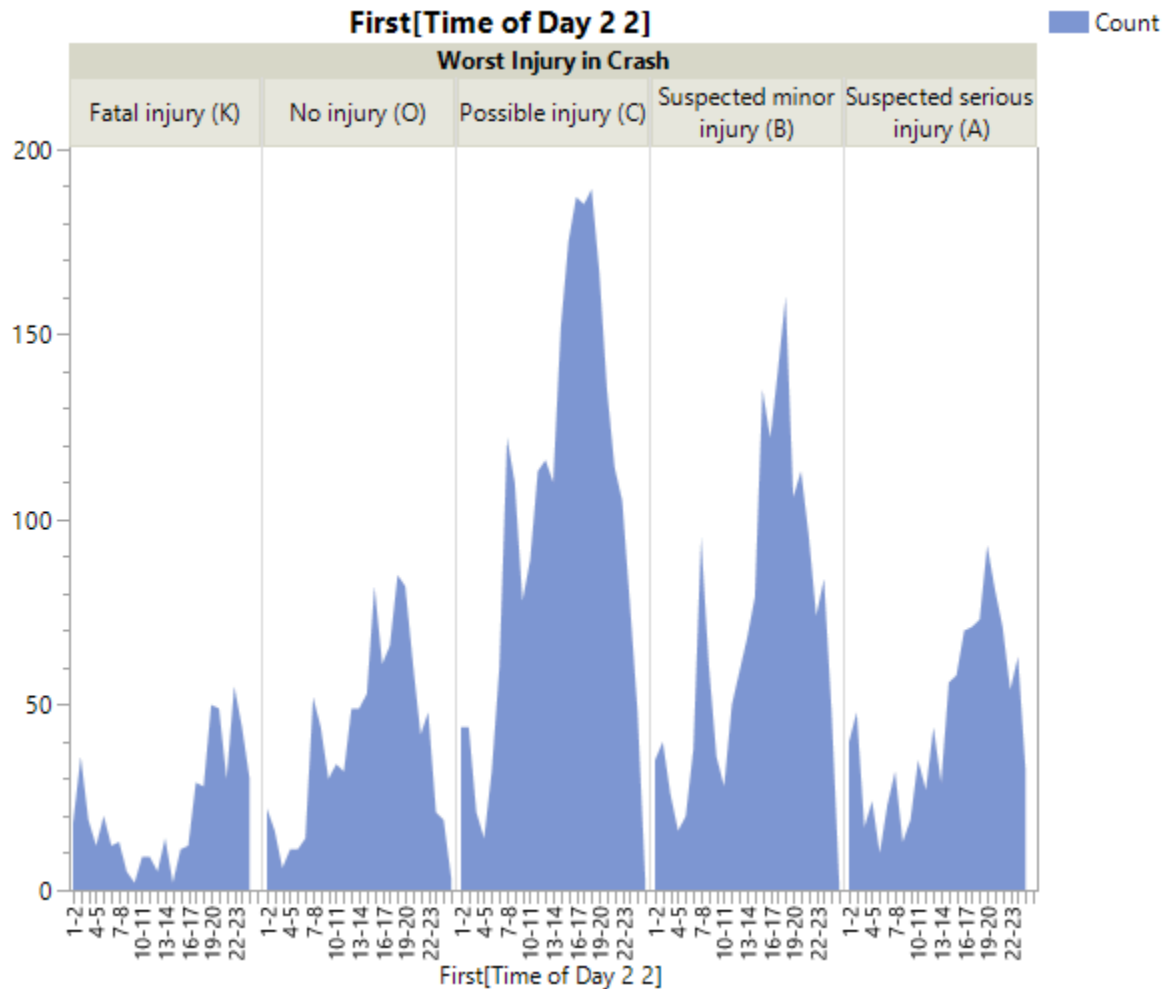
Crashes during times of day
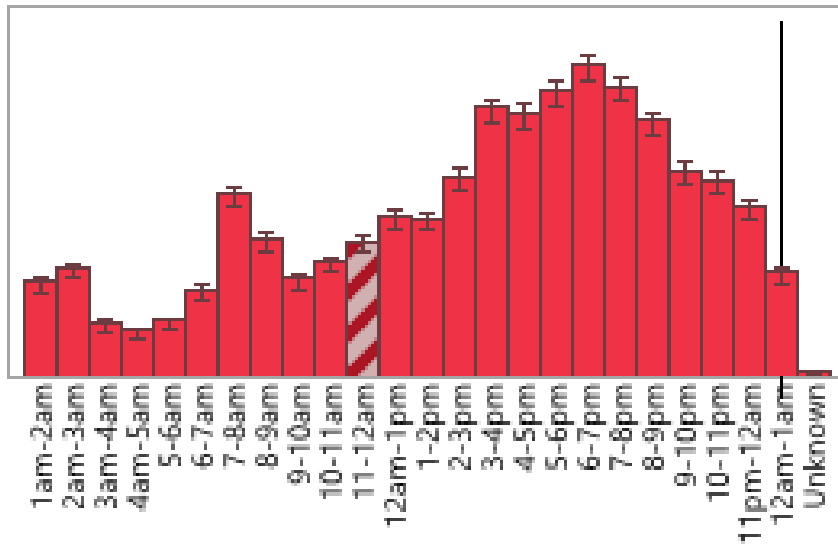
From these graphs, there appears to be some concentrations of crashes during the day.
I would like to test this hypothesis with some statistical analysis.

Statistical Analysis

Firstly, I would like to apply a logistic regression to the graph of rate of crashes over time.

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0.5615	1	1.122929	0.2893
Full	1821.5921			
Reduced	1822.1535			

RSquare (U)	0.0003
AICc	3647.19
BIC	3660.84
Observations (or Sum Wgts)	6809

Upon running the test, I return a nearly straight line with R^2 value of 0.0003. This information shows there is little relationship between fatal or serious crashes over the years. This begs the question, despite all the safety improvements made since 2010, why hasn't there been a significant decrease in the rate of serious auto-pedestrian crashes?

To help reduce auto-pedestrian crashes, I also compared the concentration of crashes throughout the day and applied a Chi-Squared Test on the distribution of crashes.
a value=0.05

Test	ChiSquare	DF	Prob>Chisq
Likelihood Ratio	276.7598	23	<.0001*
Pearson	283.0435	23	<.0001*

Null hypothesis: the proportion of fatal injury during a car crash is independent of the time of day of the car crash
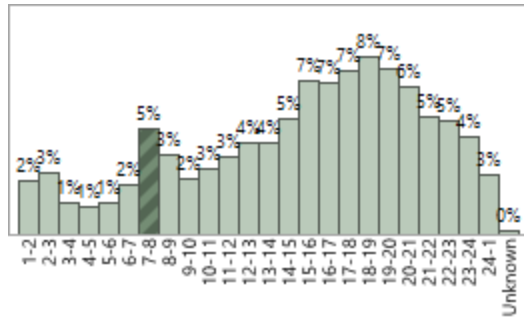H_a: The proportion of fatal injury during a car crash is dependent on the time of the car crash

Since p value is greater than 0.05, there is sufficient evidence to reject the null hypothesis. We can conclude that The proportion of fatal injury during a car crash is dependent on the time of the car crash

Fatal injuries
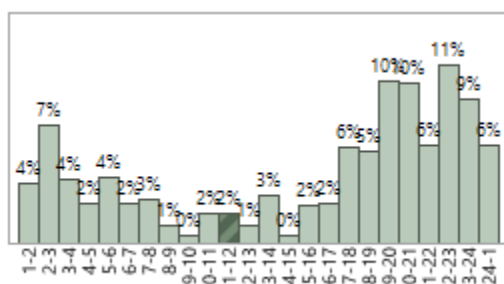
All injuries

As we can observe from the graph, the rate of crashes have a negative skew and concentrate towards early in the night. Additionally, the rate of fatal injuries appears to be bimodal and also peaks at later in the night around midnight as well as early in the morning. Both distributions demonstrate a higher likelihood of a pedestrian accident at night, when there is less adequate lighting for drivers to see pedestrians.

Conclusion and Recommendations

Based on the statistical analysis, I can conclude that there has not been a significant change in auto-pedestrian crashes since 2010 despite new innovations to improve car safety. However, there is evidence to suggest that the accidents tend to occur at night when there is lower visibility and therefore increased improvements in pedestrian detection in low lighting would help contribute to lowering the rates of auto-pedestrian accidents.

Reflection:

I was not expecting the rate of pedestrian accidents to change as little as it did over the past years since it seems like there are groundbreaking improvements made on vehicles like autopilot and obstacle avoidance every year. Some problems I ran into was having trouble running a Chi-Square Test on two crash distributions at once.

In further studies, I would like to explore other variables such as speed limit to further explore the reasons for pedestrian car crashes over the year and also extend my analysis to countries outside the US.