What Causes Traffic Fatality and How Could We Survive

Wenyi Lin

October 23, 2016

Introduction

Every day, nearly 100 people die from vehicle-related accidents. From 2010 to 2015, there are 2912 people died in 2699 fatal traffic accidents in Maryland. Basically, traffic accidents are unavoidable, however, it is possible that we face the fact and protect ourselves. Thus, we would like to know what factors could have impact on a severe accident which might cause fatality. And also, if a possibaly fatal accident happens, do we have any chance to prolong our lives, and even, save our lives.

Fatal risk could include personal chracters (age, sex, position), time factors (season, time of the day), driving behaviours (drunk, safety measurement) and environmental dectription (location). Each of these factors could be a potential threat to our safety and they could also work interactively. Here we fit a logistic model to figure out their causality and relationship.

Once a serious traffic accident happens, the response and rescue from hospitals would be the last hope to save the live. Response time and arriving time at the scene would be the first concern to construct a survival model.

Data Description

Data was compiled using the 2010-2015 Fatality Analysis Reporting System (FARS) data. We use Maryland fatality data extracted from the big data set to analyze risks in fatal traffic accidents within this state. Missing data related with age, death time was eliminated from analysis. Due to large volumn of missing data with medical service time records, thus, we applied survival analysis on nationwide dataset and used available medical sevice time variables.

Statistical Methods

1. Logistic Model

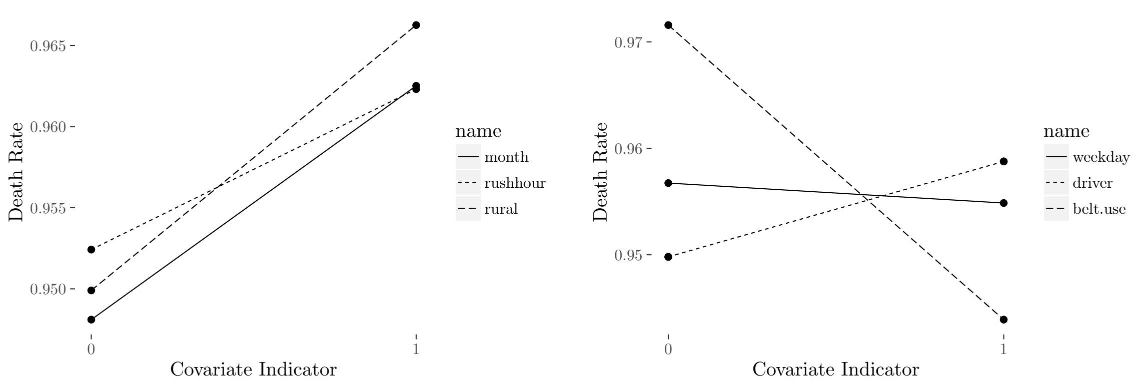


Figure 1: Possible Risk

We made plots to show the effect of possible risk factors. Basically, death rate is higher when the accident is happened at rush hour (7-10 am, 4-7pm) on weekdays between May to October. Accidents happened at rural area are likely to have dead case compared with that at urban area. Drivers are relatively to be treatened by death among all accident-relevant types of people(passenger, pedestrian, Bicyclist, other) and in fact, almost 99.5% of fatal cases happened in the car. And apparently, using seat belt could decrease the death probability. Based on our findings, we would like to consider cases only include over-16-year individuals to confirm the possibility of being a driver.

Logistic model is favorable when analyzing binary data. Based on our preliminary data analysis, we could construct a simple logistic model.

$$\begin{aligned} logit(P(Death = 1)) = \beta_0 + \beta_1 sex + \beta_2 age + \beta_3 weekday + \beta_4 month + \beta_5 rushhour + \beta_6 rural \\ + \beta_7 is.driver + \beta_8 belt.use \end{aligned}$$

Natural spline was used on age factor in order to include possible trend and the degree of freedom is 3. By cross validation, we get the AUC for simple model is 0.635.

Then we add more interactive items to the GLM model. The fatality rate in an accident would be higher is the driver is drunk. And apparently, seat belt use is highly related with driver's safety. According to these facts, our updated logistic model is given as:

$$\begin{aligned} logit(P(Death = 1)) = \beta_0 + \beta_1 sex + \beta_2 age + \beta_3 weekday + \beta_4 month + \beta_5 rushhour + \beta_6 rural \\ + \beta_7 is.driver + \beta_8 belt.use + \beta_9 belt.use * is.driver \\ + \beta_9 drunk * is.driver \end{aligned}$$

In this model, we get AUC=0.662 with cross vationdation. The table shows our estimation of model coefficients.

predictor	$\hat{\beta}$	Std (with)
Is_driver	0.683	0.306
Rural	0.409	0.159
Rush Hour	1.074	0.404
Weekday	0.067	0.164
Mid-Month	0.324	0.14
Is_driver: Belt Use	-0.3	0.338
Rush Hour:Weekday	-1.08	0.442

Thus, we conclude that compared with other person type involved in a fatal accident, the odds ratio of a driver's death is multiplied by 1.98 (1.087,3.607). Odds ratio of death at rural area is 1.505 (1.102,2.056) of that at urban area. During everyday's rush hour, odds ratio is increased by 1.927 (0.326,5.461). From the mid-Spring to mid-Autumn, odds ratio of death is 1.383 (1.051,1.819) of other seasons in a year. If a driver uses seat belt, the risk of death would be decreased by 0.259 (0.618,-0.437).

Actytually, the analysis is umcomplete to some extent. Fisrt of all, it is better to include all accident data over 5 years in Maryland to make inference about fatality rate. Secondly, more factors could be included in the model to improve the accuracy of the predictive model.

2. Survival Analysis

A fatal accident could always happen no matter how we try to avoid it. In this situation, it is critical porlong the survival time or even save lifes. It is believed that medical response time could affect survival time after traffic accidents.

Notificaiton time is the time we used to inform the emergency medical service after the accident happens. The time duration that emergency medical service arrived on the crash sceneEmergency medical service arrived on the crash scene is called arrival time. Both of them could be calculated based on FARS data.

Figure 2A Fatal Accident Survival Time

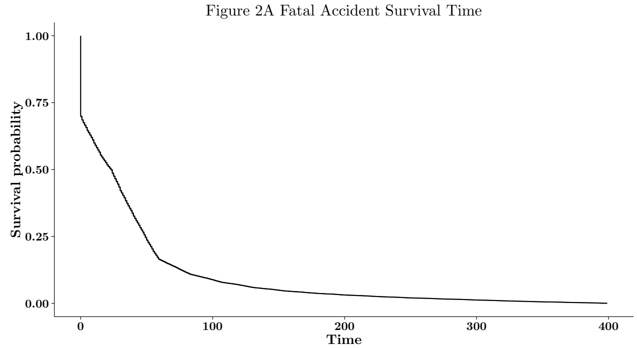
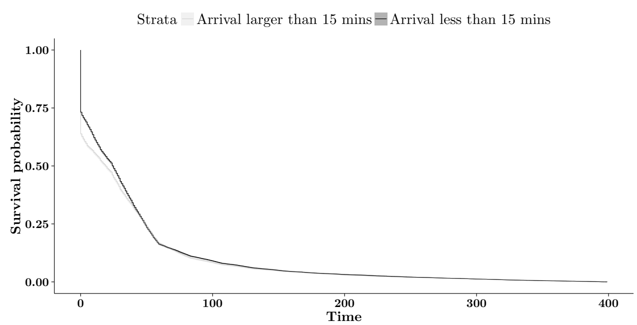


Figure 2B Fatal Accident Survival Time Stratified by Arrival Time



Lag time is the period between the time of the crash and this person's time of death. Figure 2A shows the fatalities over time. This fatality curve, or survival probability curve over time, S(t), indicates that approximately 0.774 of fatalities occurring within 6 hours occurred within 100 minutes of the crash (including 0.331 instant deaths before time = 10). Then, we plot fatality curve for stratified dataset based on notification time and arrival time. Figure 2B indicates that early arrival time (arrival time < 10) could help with decrease the proportion of instant death.

Based on this, we construct a Cox Proportional Hazard Model (Cox-PHM). Actually, emergency medical service often based on the location of the crash scene. Thus, we include rural indicator in our model:

$$\mathrm{h}(t) = h_0 \exp(\beta_1 Arrival + \beta_2 Location)$$

The model included two factors treated as binary variables: arrival time (later than 10 minutes versus earlier arrival) and crash location (rural versus urban). From the results, we conclude that later arrival time could carry a 2.14% (0.38%,3.92%) relatively higher fatality hazard than earlier group.

Now we could say that earlier arrival of emergency service helped with increase of survival probability since most of the fatal case happens very quickly. However, this simple survival analysis could not include all of the imformation to estimate survival rate. And unfortunately, the time of death after injury and emergency service arrival may be inconsistently recorded. A thorough analysis of the data would required if we want a more precise result.

Conclution

As for fatal traffic cases, many factors would influence the death of a person involved. Generally speaking, drivers in the accident are the most vulnerable people and the situation could be even worse if they don't wear a seat belt. The fatality rate would increase when roads get crowded, such as rush hour, week day, and relatively warm seasons. And also, in-time emergency medical service could help to decrease the probability of early mortality and accordingly, create chances to save lives.

Reference


```
## [1] M. Rosekind and D. Patil. _2015 Traffic Fatalities Data Has
## Just Been Released: A Call to Action to Download and Analyze_.
## 2016. <URL:
## https://www.transportation.gov/fastlane/2015-traffic-fatalities-data-has-just-been-released-call-acti
##
## [2] M. C. Wu Jingshu and A. Longthorne. "Updated Estimates of
## Potential Traffic Fatality Reductions With Automatic Collision
## Notification". In: _24th International Technical Conference on the
## Enhanced Safety of Vehicles (ESV). No. 15-0269._ (2015).
```