U.S. Roadways

Michael Jadoo[1]
[1]Joint Statistical Meetings, Chicago, Illinois

Abstract

Every year many lives are either lost or life changing injuries are experienced that could have been prevented while driving on the U.S. roadways. With the use of the National Highway Traffic Safety Administration National Automotive Sampling System General Estimates System databases, an analysis was conducted to reveal which areas in the United States is the deadliest place to drive and what factors predicate the occurrence of accidents. The databases were used to conduct a statistical analysis on the maximum injury severity variable and other associated variables in the dataset, over time. The report presents the state and area that is prone to accidents and what are the contributing factors of these crashes. As such, the report offers a practical framework to preventing accidents by analyzing what predicates them.

Introduction

Undeservedly, preventable road accidents cause premature death and debilitating life altering injuries on U.S. roadways. In with aim of highlighting accident prone areas and developing a meaningful framework to prevent accidents, a statistical analysis was conducted using the National Highway Traffic Safety Administration (NHTSA), National Automotive Sampling System (NASS) General Estimates System (GES) database which sought to answer the following research questions.

(1) Which areas have a high prevalence of fatal accidents?

(2) What external conditions predicate accidents?

(3) What internal conditions contributed to fatal crashes in the south?

(4) What is the demographic information about the individuals who were involved in fatal crashes in the South region?

(5) Which areas is the safest place to be in your car?

(6) Were the safety devises working?

(7) Which car is mostly associated in fatal accidents?

The purpose of the report is to first highlight the accident prone areas. Subsequently, analyze the highlighted areas in order to clearly unmask what is contributing to accidents and who is most affected. To accomplish these tasks two additional empirical analyses was done using the Fatality Analysis Reporting System to identify the state with the highest frequency of fatal crashes and to ascertain whether high population was a contributory factor in increase of accidents.

Data Collection

An empirical analysis was conducted using crash data files from 2009 to 2014 from the NASS GES databases. The data centers were established to identify problems related to traffic safety problem areas with the aim of aiding in the formulation of regulatory initiatives for road users as well as analyze the effectiveness of implemented interventions. According to the Analytical User's Manual 1988-2014, the NASS GES database is nationally representative of a probability sample coming from more than five million police-reported crashes within a given year. The data is collected from approximately 400 police agencies at 60 sites, and gathering 50,000 randomly sampled police accident reports (PAR) per year.

For a crash to be qualified for the NASS GES sample a PAR must be completed. This requires that the accident must involve at least one moving vehicle traveling on a roadway that resulted in damage to property, injury, or fatality. The data collected provides annual estimates for four geographical U.S. regions: the Northeast, Midwest, West, and the South, which match the Census Bureau's regional designation. The report utilized 10 common datasets from NASS GES in each year from 2009 to 2014 in order to perform the analysis. The data sets were ACCIDENT, EVENT, DISTRACT, FACTOR, IMPAIR, MANEUVER, PERSON, TRAFCON, VEHICLE, and VISION tables. Subsequently, the datasets were merged together and used to collect information from several topics to help identify potential areas to focus on in order to reduce fatal crashes over time.

Moreover, the fatal crashes estimates from the Fatality Analysis Reporting System (FARS) were also used in this report in order to create choropleth maps to illustrate the fatal crashes totals and growth rate of fatal crashes by state. The FARS is a nationwide census providing the NHTSA, governmental, and public entities with a yearly report of fatal injuries that was caused by automobile crashes. The FARS provided data on fatal crashes involving all types of vehicles and is used to identify highway safety problem areas. FARS is a census of fatal motor vehicle crashes with data files documenting all fatalities that had occurred across the 50 states, District of Columbia, and Puerto Rico. In order for an accident to qualify for the database system of FARS, the crash had to involve a motor vehicle traveling on a roadway open to the public, and the accident must have resulted in death of a motorist or non-motorist within 30 days of the crash. The data table used was the fatal crashes by state estimates for 2014 and the growth rate of fatal crashes by state for 2013 to 2014 for the choropleth map. Additionally, the data series from national fatal crashes estimates were applied and covered a time period from 1994 to 2014. The FARS licensed drivers estimates obtained from the Federal Highway Administration (FHWA) database for total number of licensed drivers stratified by age and gender for the 50 states and District of Columbia. It includes learner’s permits, non-drivers identification cards, and any license in a suspended status as of the end of the reporting year.

To analyze the number of accidents and the correlation to the population density, population statistics were obtained from the U.S. Census Bureau. The dataset first provides an estimate with a population base from the decennial census, then they include births, subtract deaths, and add net migrations for both international and domestic migrates (Census).

Conceptual framework

Using the NASS GES datasets, an empirical analysis was done to answer the seven questions, to deduce the factors that may be attributing to fatal accidents. The data tables used covered the period from 2009 to 2014, which provided a trend analysis. Additionally, the FARS data tables from 1994 to 2014 were used to construct a linear model. To being with, a choropleth map was made to illustrate the areas that experience the most fatal crashes as well as provide information on the measure of population distribution and dispersion for the areas.

A linear regression model was constructed to test the hypothesis on whether the U.S. population growth is contributing to growth in fatal crashes. This situation comes from part of the findings by David Shinar in his 2007 manual Traffic Safety and Human

Behavior (2007) in which he states that global rise in population contributed to a significant rise in the vehicle accidents worldwide. Additionally, the external and internal conditions contributing to these accidents were considered. External conditions in this report were factors in the driver's environment such as, speed limit, time of the accident, and weather conditions. The internal condition factor is defined as a situation that influences the driver's judgment, so the focus was placed on alcohol use. A logistic regression analysis was performed to correlate whether day and time of driving while speeding would give a higher probability for a driver to get involved in a fatal accident. After that, a correlation was made using the compound growth rate between the number of accidents occurring by age group and gender. Additionally, the report investigated whether car related issues could have contributed to the accident.

Estimation methods and models

A linear regression model was constructed to test the hypothesis of whether the U.S. population growth, over time, has caused fatal crashes to increase. The estimate of driver license issuance was also used in the model. However, after inspecting the dependent variable and its regresses, it became apparent that there was a structural break in 2006. This situation resulted in choosing the variable estimate of total driver's license being issued to be used in the model. An assumption was made that if one gets a driver's license they are part of the U.S. population and they are trained to be able to drive on the road. Furthermore, having a driver's license means a person can have an automobile and drive on the road. Two models were made and the Chow test was constructed to test whether population increase has resulted to more accidents.

$$Log\ (Fatal) = \beta_0 + \beta_1 Div\_lic_{pre2006} + u$$

$$Log\ (Fatal) = \beta_0 + \beta_1 Div\_lic_{post2006} + u$$

$$Chow = \frac{(RSS - RSS_{pre2006} - RSS_{post2006})/k}{(RSS_{pre2006} + RSS_{post2006})/(n_{pre2006} + n_{post2006} - 2k)}$$

Additionally, a logistic regression model was constructed using the tables from 2014 NASS GES to help analyze the most recent external condition of time of fatal accidents. A logistic regression model was done on the response variable CRASH, which identifies the probability that a crash or fatal crash occurred at a specific hour. This was done using the independent variable COUNT, which is defined as the total fatal crashes in a specific area and time, HOUR_IM, the imputed hour at the crash occurred, and DAY_WEEK, which is the day of the week that the crash occurred.

The crash response variable was defined as

$$y_i \begin{cases} \textit{0- fatal accident didn't occur at a time that was above average} \\ \textit{1- fatal accident did occur at a time that was above average} \end{cases}$$

The logit model equation

$$P(y=1)=\frac{\exp(\beta_0+\beta_1 Count+\beta_2 time\ +\beta_3 day\_week\ )}{1+\exp(\beta_0+\beta_1 Count+\beta_2 time\ +\beta_3 day\_week\ )}$$

Results

The first question was, "*Which areas have a high prevalence of fatal accidents*?" Using the 2014 tables from the NASS GES database, the number of fatal accidents by region was illustrated using a choropleth map. The South region was reported has having 1,629 fatal crashes which is the highest number compared to other regions (figure 1). Moreover, from 2009 to 2014 the south was reported to have the highest frequency of fatal accidents (figure 2).

Figure 1. U.S. regional map with count of fatal crashes in 2014

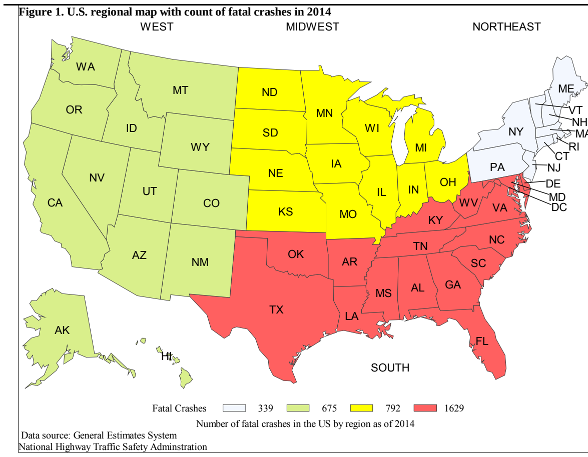
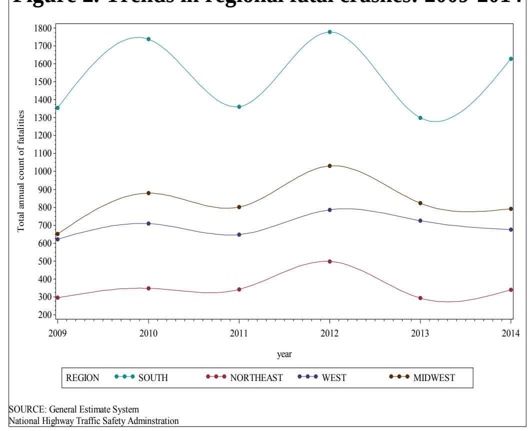


Figure 2. Trends in regional fatal crashes: 2009-2014



According to the GES, the South region has the highest area of fatal crashes. Moreover, using the FARS dataset, it shows that in 2014 Texas was the state with 3,538 fatal accidents, the most compared to other states. Additionally, table 1 shows the relative measure of central tendency, dispersion, and distribution of each region using the mean, median, coefficient of variation, skewness, and kurtosis.

Figure 3. State total of fatal crashes in the US during 2014

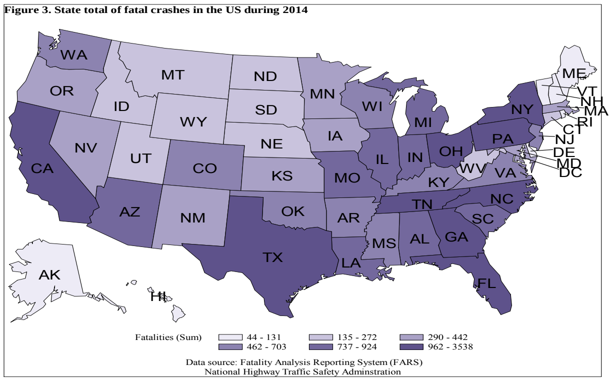


Table1			**Regional Relative Measures**		
2014	Mean	Median	Coefficient of Variation	Skewness	Kurtosis
Northeast	409.7	248	105.9	1.2	-0.1
Midwest	534.4	446	59.6	0.2	-1.6
South	935	703	92.5	2.2	5
West	521.2	290	151.6	3.3	11.1

The coefficient of variation shows that the Northeast, West, and South regions have a high degree of dispersion of fatal crashes across their states. Conversely, the Midwest region shows fatal accidents to be less dispersed and the reported mean and median reveals that this region has approximately the same level of fatal crashes being reported across its states. The skewness measure reports that the South, West, and Northeast are positively skewed which means that these areas have few states that are experiencing extreme levels of fatal crashes compared to its neighbors. The kurtosis shows values that describes the regions was heavy tailed which signals extreme levels (outliers) of fatal crashes in some states compared to others.

Table 2: Annual Growth rate of Fatal crashes and population

Year	Fatal Crashes	Civilian Population growth
2005	2.1	0.9
2006	-1.5	1.0
2007	-3.1	1.0
2008	-8.7	1.0
2009	-9.7	0.9
2010	-1.8	0.8
2011	-1.4	0.8
2012	3.8	0.8
2013	-2.6	0.8
2014	-0.7	0.7
Source:	FARS, U.S. Census	

The US population growth as well as its issuance of driver's licenses has increased (since 1994) over the years. However, according to the FARS dataset it appears that since 2006 fatal accident have decreased significantly ranging from -1.5, in 2006, to -9.7, in 2009. From this finding it appears that increased population growth did not contribute to the rate of motor vehicle accidents as it was stated by David Shinar in his 2007 manual Traffic Safety and Human Behavior.

Within the 1994 to 2014 period an empirical analysis was done to determine the hypothesis whether an increase in national population contributes to fatal crashes in the U.S. is true. Using the FARS fatal crashes estimate, U.S. Census civilian population estimates, and the Federal Highway Administration estimate of drivers licensed issued appears that fatal crashes have decreased over time (table 3).

Table 3. Linear Regression estimates	**Model 1**	**Model 2**	**Model 3**
Period(s):	1994-2014	1994-2006	2006-2014
	Log(Fatal)	Log(Fatal)	Log(Fatal)
Driver License (thousands)	-0.00000544 (0.00000134)	0.00000215 (0.00)	-0.00002676 (0.00000374)
Constant	11.55 (0.26442)	10.13 (0.06)	15.99 (0.78)
Observations	21	13	9
R2	.465	.80	.88
Rooted Mean Squared Error (RMSE)	.0763	.01	.03
Akaike information criterion (AIC)	-106.15	-117.74	-57.45
Notes:	-The civilian population estimate was omitted in the all the models due to its near prefect collinearity with drivers licenses variable. -The fatal variable was normally distributed for the 1994 to 2006 time period; however, for model 2 the natural log was used in order to compare this model to its counterparts.		

After visual inspection of the series, the amount of fatal accidents occurring from 1994 to 2014 experienced a structural break in 2006. In the 1994 to 2014 model, for each additional drivers licensed issued the percentage change of fatal crashes decrease by -0.54.

During 1994 to 2006, the issuances of driver's licenses, per person, actually had a positive percentage growth of 0.215. This observation means that during 1994 to 2006 as more residents in the population attained drives licenses this contributed to more fatal accidents.

However, the empirical analysis covering 2006 to 2014 shows that as more people got drivers licenses this resulted in a negative growth of -.27 for every person given a license to drive on the road.

For each of the models constructed the estimates of drivers licenses being issued had an 80% explanatory power on the variation of fatal accidents occurring in 1994 to 2006 and 88% explanatory power for 2006 to 2014. For the whole period being observed, 1994 to 2014, the estimates of driver's licenses being issued has a 45% explanatory power on the variation of fatal accidents occurring.

The chow test was done to check to see if this structural break in 2006 was valid using a level of significance of 0.05.

$$Chow = \frac{(.11 - .001 - .009)/2}{(.001 + .009)/(21 - 2*(2))} = 85$$

4.45 < 85
Null hypothesis rejected

With F (2, 17) degrees of freedom, the critical value from the F-table is 4.45. This resulted in rejecting the null hypothesis of no difference between the periods of 1994 to 2006 and 2006 to 2014.

So it was concluded that something happened in 2006 which caused fatal crashes to occur less frequently. According to a report by Dr. Dischinger and others (2013), despite the increase in population and roadway traffic, in 2006 many factors contributed to the decline in fatal crashes. Those factors were improved changes in roadway design, campaigns to reduce drunk driving and increase seatbelt use, and improved vehicle safety standard for crashworthiness.

Furthermore, in 2010, an analysis by Longthorne, Subramanian, and Chen stated that due to the recession period between 2008 to 2009 high unemployment rate had coincided with the large drop in fatal crashes during the same period. This phenomenon was also apparent during the recessionary period of 1980s and early 1990s (Longthorne). One could conclude that because drivers have less money or the inability to spend more for recreation activities less people would be on the road. They would find other activities less costly to participate in. In summary, the hypothesis on whether increased U.S. population growth contributed to the increase of fatal crashes is false. It appears that in each geographical area the frequency of fatal crashes would be different. Returning to answer of the first question, the NASS GES reports that the South region is the area with the most fatal crashes being experienced in 2014 with the Texas being the outlier state.

The second question was, "*What external conditions predicate accidents*?" The external conditions that were analyzed were traveling speed, time of day, and weather. Using the NASS GES database, the contributing external factors that lead to fatal accidents during 2014 in the South was; traveling over the speed limit, driving during rainy weather conditions, traveling during rush hour, and riding on the road during periods for weekly festive activities. Weekly festive activities is defined as going to different recreational establishments to associate with friends within a particular area during evening (approximately 8:00 pm to 9:00 pm) to early morning periods (approximately 1:00 am to 4:00 am) from Thursday thru Saturday. Rush hour is defined as times when the population travels on the road to go to work during the weekday (Monday-Friday) between the hours of 7am to 9am in the morning and 3pm to 7pm in the afternoon.

An empirical analysis was conducted to show what day of the week and time does fatal accidents or accidents generally occur. The GES data indicated that most fatal crashes in the South occur during rush hours periods. However, there are exceptions in which there are a large amount of fatal crashes occurring during non rush hour times. This could be due to a different groups within the work force that have occupations that require them to work late at night which results in traveling during non-rush hour periods (Appendix B).

There are some weekdays that experience higher frequency of fatal accidents is occurring by 11am to noon (lunch time) this phenomenon is apparent on Wednesdays, Thursdays, and Fridays. During the evening and weekend, for recreational activities, different but expected finding were discovered. It appears that fatal accidents starts to occur on Thursdays by 8:00 pm to 11:00 pm possibly due to the typical college night outing. On Fridays, fatal accidents start occurring around 9:00 pm to about 4:00 am on Saturday morning. Then on Saturday evening starting around 8:00 pm fatal accidents starts to occur up until 5:00 am on Sunday morning (Appendix B).

A correlation analysis was done to see if driver over the posted speed limit that causes fatal accidents and just fatal accidents occurring during each day of the week were related to each other (table 4).

Table 4: Time of fatal crash while speeding & time of fatal crash occurring

Day of the week	Correlation Coefficient
Sunday	0.94
Monday	0.88
Tuesday	0.82
Wednesday	0.89
Thursday	0.89
Friday	0.75
Saturday	0.85

The result of the analysis shows that for each day of the week the correlation coefficient s were between 74% to 95%, which indicates that there is a moderate to strong positive linear relationship between speeding and fatal accidents occurring during the week. The interpretation is when a driver speeds during rush hour or recreational activities they have a strong chance of getting involved in a fatal accident.

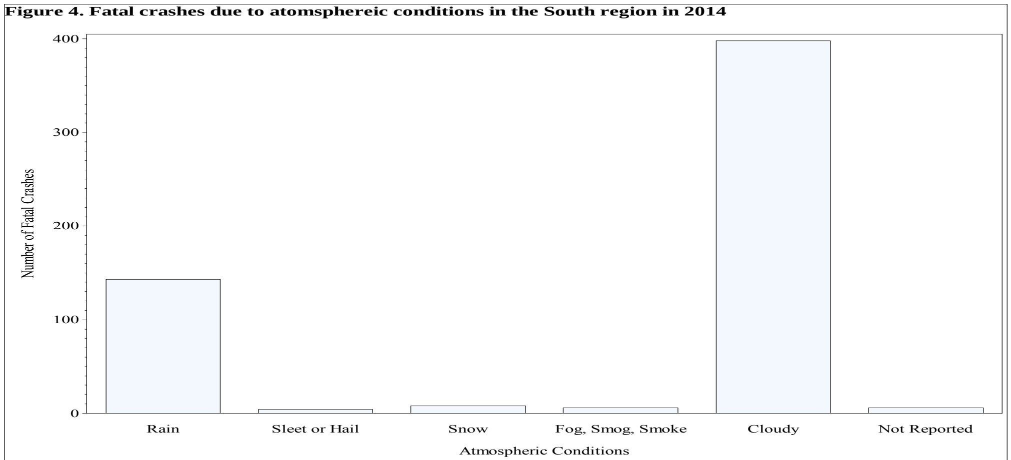
Two supplemental logistic regressions were constructed to determine the probability of getting into a fatal accident while speeding because of time and day of the week. The logit model showed the variables hour and day of the fatal accident, as well as the number of fatal crashes explain 69% of the variation of the probability in the likelihood that one would get into a fatal crash while speeding. The probit model reports its models explanatory power to be at 68%. However, both models report the percentage of correctly predicted estimates to be at 99.3. The logit model was selected to determine the probability of getting into a fatal crash while speeding.

This model reports that the odd of getting involved in a fatal accident if you are speeding giving the above average frequency of fatal accidents, in the South region, is 35% in 2014 in the total number of fatal crashes and at a 95% level of confidence this increase ranges between 17.9% to 55%.

The odds of getting involved in a fatal accident if you are speeding giving the above average frequency of fatal accidents in the South region is 70% in 2014 for each day in the week and 90% for the time during the day. However, the day during the week and time during the day showed to be not significant (Appendix C).

For atmospheric conditions, the GES reports that in 2014 rainy weather contributed to fatal crashes; however, most fatal crashes occurred during dry conditions as seen in figure 4 below.

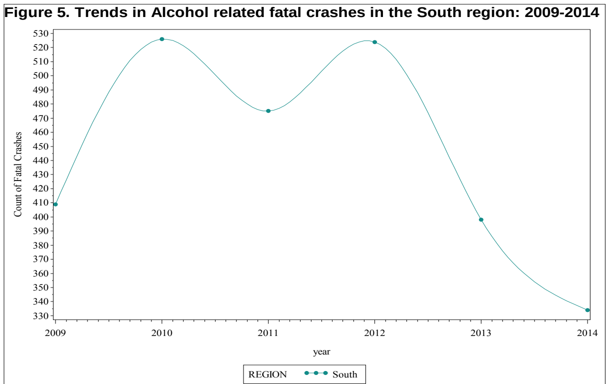
Figure 4. Fatal crashes due to atomsphereic conditions in the South region in 2014



SOURCE: General Estimate System
National Highway Traffic Safety Adminstration

The third question was, *"What internal conditions contributed to fatal crashes in the south?"* The Internal condition was defined as alcohol involvement causing a fatal crash. The NASS GES manual defines alcohol involvement as the alcohol use involved in a crash as reported by the police.

Figure 5. Trends in Alcohol related fatal crashes in the South region: 2009-2014



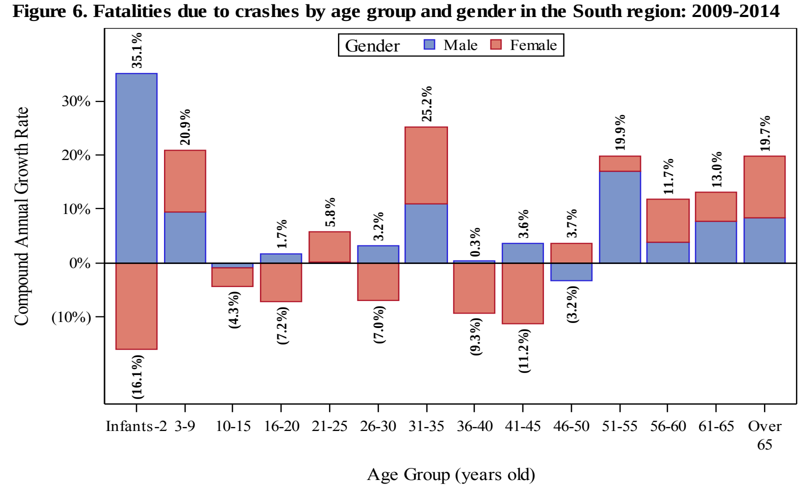
SOURCE: General Estimate
System National Highway Traffic Safety Adminstration

Surprisingly this measure's, which resulted in fatal crashes, trend has decreased significantly as seen in figure 5. This chart shows that crashes involving drinking are contracting significantly.

Moreover, this finding coincides with Dr. Longthorne and others report that during periods of recession, as it was for 2009, individuals will forego opportunities to engage in festive activities due to budgetary constraints (Longthorne, et al. 2010).

The forth question was, *"What demographic information about those involved in fatal crashes in the South region."* Using the imputed age and gender variables, a comparison was done using age groups of both male and female to show demographic information derived from the NASS GES datasets. The compound annual growth rate was used to construct the average growth rate from 2009 to 2014. This analysis showed some surprising results.

Figure 6. Fatalities due to crashes by age group and gender in the South region: 2009-2014



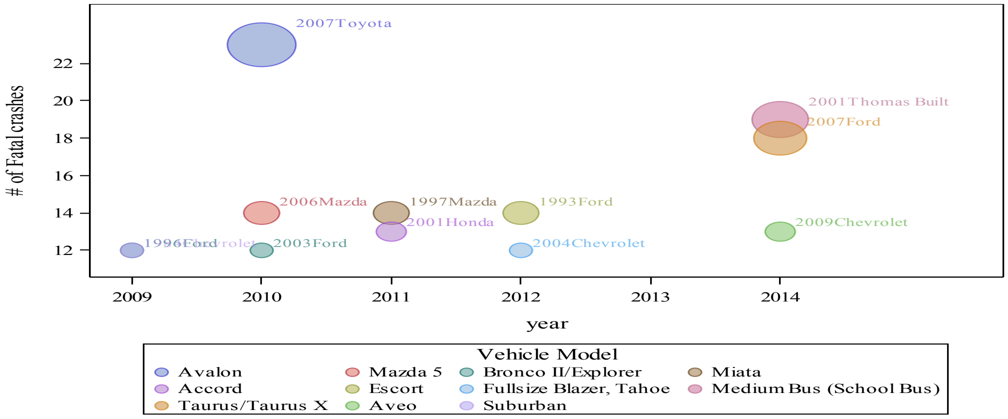
SOURCE: General Estimate System
National Highway Traffic Safety Adminstration

Figure 6 shows that, in the South region, infants to two year olds males have the highest rate of fatalities due to crashes with a growth of 35.1 percent. Conversely, girls in the same age group have the lowest with a decline of -16.1 percent. Women between the ages of 31 to 35 have the highest rate of fatal crashes for their gender with 25.2 percent increase. Males between the ages of 46 to 50 have the lowest during this period with a contraction of -3.2 percent points.

The fifth and sixth question was, *"What is the safest place to be in your vehicle and are safety devise working."* The GES reports that in 2014 the backseat location in the automobile may provide the safest location to be in during an accident, as most fatalities are occurring at the driver's seat location during a crash. Fortunately, seatbelts and airbags are working to prevent fatalities.

The seventh and final question was, *"What type of car is mostly associated in fatal accidents."* Constructing an empirical analysis during 2009 to 2014 for type of automobile mostly associated in fatal accidents within a given year shows, as seen in figure 7 below, that in 2010 the 2007 Toyoda Avalon had the highest fatal crash amount against other make and models. In 2014, 2007 Ford Taurus had the highest fatal crash total along with the 2001 Thomas Built School bus.

Figure 7. Make and model of vehicle with large fatality count from 2009-2014 South region



SOURCE: General Estimate System
National Highway Traffic Safety Adminstration

Conclusion

In summary, the NASS GES for regional level and the FARS tables for individual's state level crash estimates for 2014 revealed several interesting findings. The results showed the South region experienced the highest estimate of fatal crashes in the nation with an estimate of 1,629. The state of Texas was reported as having an extreme amount of fatal crashes compared to other states that year with an estimate of 3,538. The FARS and FHWTA datasets revealed the downward trend in fatal accidents. The level of fatalities due to vehicle accidents has contracted compare to 1994. Additionally, in 2006 this contraction started its decent due to several factors; improved changes in roadway design, campaigns to reduce drunk driving and increase seatbelt use, and improved vehicle safety standard for crashworthiness (Dischinger et al, 2013).

Moreover, the recessionary period from 2008 to 2009 is also believed to contribute to this downward trend due to less disposable income available for drivers wishing to engage in festive activities (Longthorne, and other 2010). These findings disprove the earlier hypothesis on whether an increase population growth within the context of a specific country resulted in an increase growth of fatal crashes.

In 2014, the contributing external for fatal crashes were going over the posted speed limit, driving during rainy conditions, traveling during rush hour, and riding on the road during periods for weekly festive activities. On a more positive note, the internal factor of fatal crashes involving alcohol has decreased significantly. The safest location in the vehicle was reported to be at the back seat and the 2007 Ford Taurus is reported to be the vehicle that is related to the most frequent fatal crash. Women between the ages of 31 to 35 have the highest rate of fatality for their gender as well as male infants to two-year-olds.

Some recommendations after the findings were discovered would be in the form of media outreach in the form of TV or online commercial programming that targets parents in regard to child safety seats and encouragement to travel less on the road with children.

Additionally, social media could be used to effectively reach out to everyone to practice safer driving habits. Moreover, other forms of work schedules should be incorporated in the private sector to keep drivers off the road during rush hour periods.

There is room for further research from this report as one can find other variables that contributes to fatal accident. Additional, one could look at road infrastructure related causes as well as devices that could cause a person to be distracted while driving. With these and other additional findings we can come to a more accurate assessment on contributing factors to fatal crashes across the nation.

Acknowledgements

I would like to thank the staff of the National Highway Traffic Safety Administration for their data support.

References

Census. METHODOLOGY FOR THE UNITED STATES POPULATION ESTIMATES: VINTAGE 2015. Washington, DC: U.S. Census Bureau. <https://www.census.gov/popest/methodology/2015-natstcopr-meth.pdf>

Geography. *Geographic Terms and Concepts - Census Divisions and Census Region.* Washington, DC: U.S. Census Bureau. < https://www.census.gov/geo/reference/gtc/gtc_census_divreg.html>

GES. National Highway Traffic Safety Administration. *National Automotive Sampling System General Estimate System Users Manuel 1988-2014*. Washington, DC: National Center for Statistics and Analysis, DOT, 2015. Print.

GES. National Highway Traffic Safety Administration. *Database.* June 8, 2016 <ftp://ftp.nhtsa.dot.gov/GES/>

FARS. National Highway Traffic Safety Administration. FARS Analytical User's Manual 1975-2014. Washington: National Center for Statistics and Analysis, DOT, 2015. Print.

FARS. National Highway Traffic Safety Administration. *Database FARS Encyclopedia*. June 8, 2016 <http://www-fars.nhtsa.dot.gov/Main/index.aspx>.

FHWA. *Highway Finance Data Collection*. Washington, DC: Federal Highway Administration.
< http://www.fhwa.dot.gov/policyinformation/hss/guide/ch4.cfm >.

McGrew, Lembo, Monroe. An Introduction to Statistical Problem Solving in Geography, 3rd Ed. Descriptive Problem Solving in Geography page 44-47

Shinar, David. Traffic safety and human behavior. Volume 5620. Emerald Group Publishing Limited; 1 edition. 2007. Page 5.

Dischinger, Patricia. Ryb, Gabriel. Kufera, Joseph. Ho, Shiu. (September 2013). *Declining Statewide Trends in Motor Vehicle Crashes and Injury-Related Hospital Admissions.* PMCID: PMC3861824. PubMed Central, US National Library of Medicine NIH. Chicago, Il: Association for the Advancement of Automotive Medicine. < http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861824/>

Longthorne, Anders. Subramanian, Rajesh. Chen, Chou-Lin. "An Analysis of the Significant Decline in Motor Vehicle Traffic Crashes in 2008". June 2010. DOT HS 811 346.Mathematical Analysis Division, National Center for Statistics and Analysis. Washington, DC: National Highway

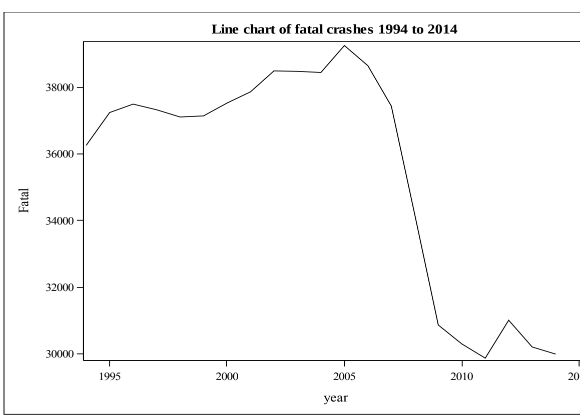
APPENDEX A

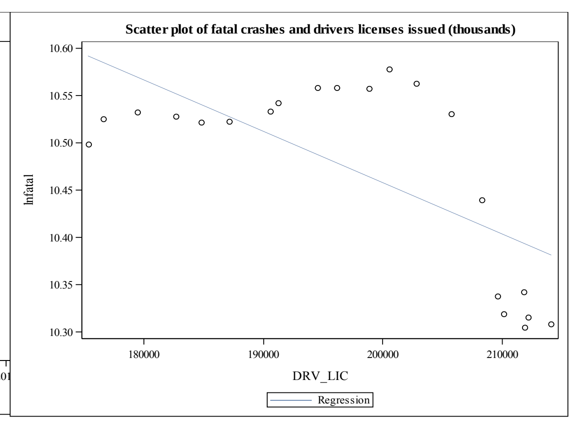
Linear model showing structural break.

SECTION 1 Variable list

Fatal	Number of fatal crashes in a given year
Drv_lic	Number of driver licenses issued in a given year in thousands
POP	Estimated civilian population within a given year in thousands

SECTION 2





SECTION 3

Descriptive statistics

Variable	OBS	Mean	Std Dev.	Min	Max
Fatal	21	35480.48	3469.65	29867	39252
DRV_LIC	21	197379.29	12766.79	175403	214092
POP	21	291108.62	18921.77	260327	318857

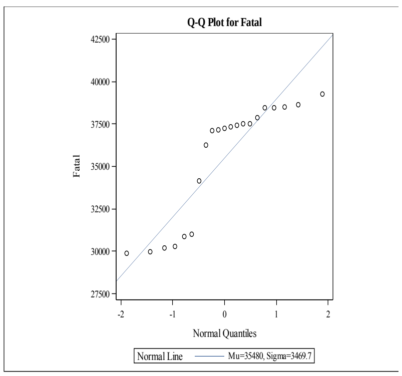
SECTION 4

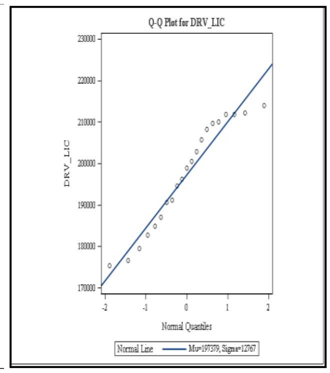
Correlation Coefficients 1994 to 2014

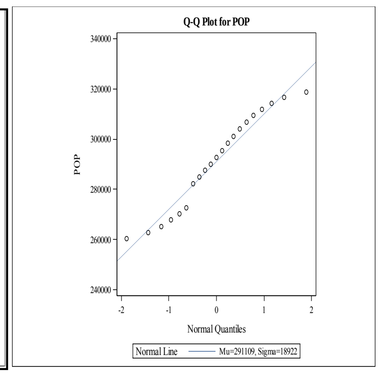
Table 1	Fatal	DRV_LIC	POP
Fatal	1	-0.67 (0.0008)	-0.688 (0.0006)
DRV_LIC	-0.67 (0.0008)	1	0.9939 (<.0001)
POP	-0.688 (0.0006)	0.9939 (<.0001)	1

SECTION 5

Normality-QQ PLOTS







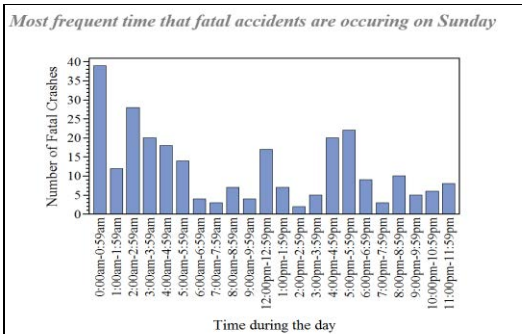
SECTION 6

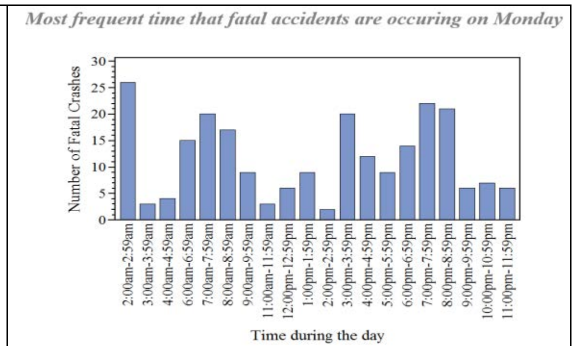
Structural break at year 2006.

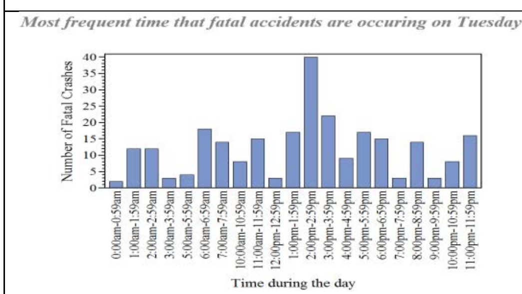
Table 2. Linear Regression estimates	Model 1	Model 2	Model 3
Period(s):	1994-2014	1994-2006	2006-2014
	Log(Fatal)	Log(Fatal)	Log(Fatal)
Driver License (thousands)	-0.00000544*** (0.00000134)	0.00000215*** (0.00)	-0.00002676*** (0.00000374)
Constant	11.55*** (0.26442)	10.13*** (0.06)	15.99*** (0.78)
Observations	21	13	9
R^2	.465	.80	.88
Rooted Mean Squared Error (RMSE)	.0763	.01	.03

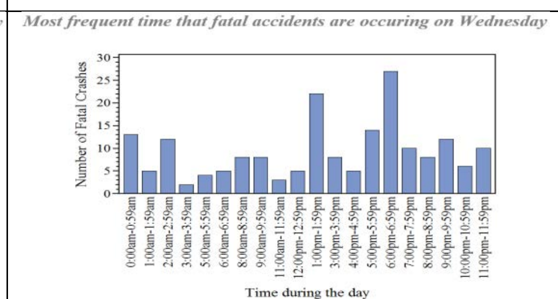
Akaike information criterion (AIC)	-106.15	-117.74	-57.45
Notes:	The civilian population estimate was omitted in the all the models due to its near prefect collinearity with drivers licenses variable. The fatal variable was normally distributed for the 1994 to 2006 time period; however, for model 2 the natural log was used in order to compare this model to its counterparts. *=p<0.1, ** = p<0.05, *** = p<0.01		

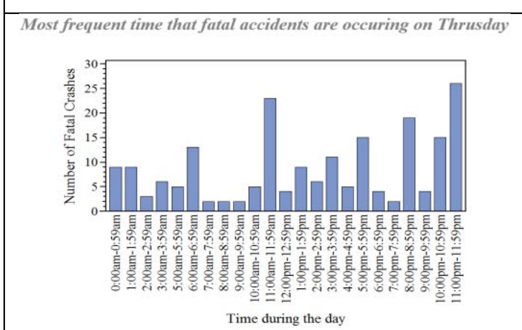
APPENDEX B

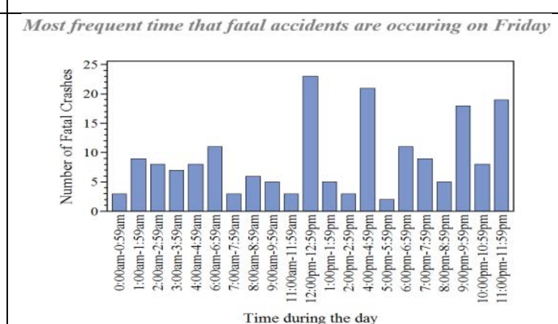


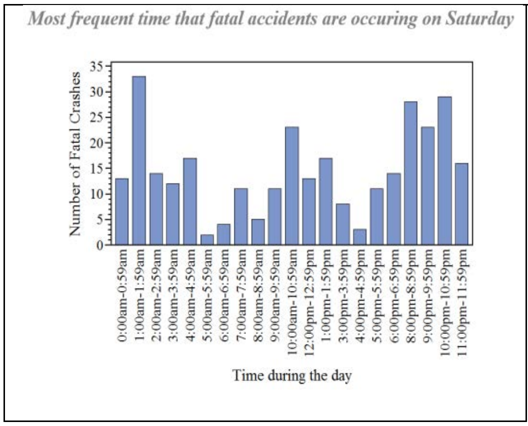












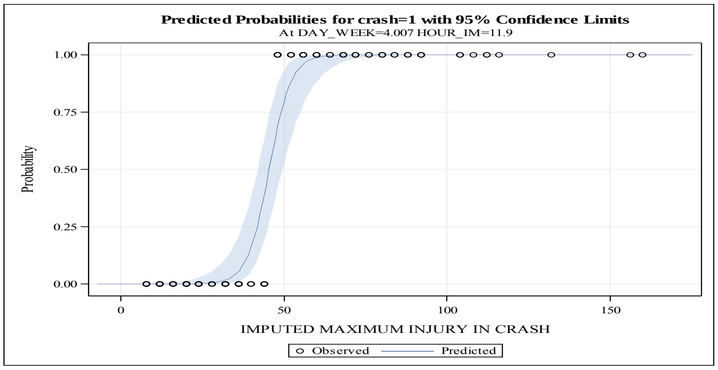
APPENDEX C

Logistics model for probability of getting a fatal accident because of speeding during time of day

Section 1: Variable list

Crash (binary response)	If a fatal accident occurs because of speeding that is above average
Hour_im	Imputed time (hour) that a fatal accident occurs
Day_week	The day during the week that a fatal accident occurs
Count	The total amount of fatal crashes occurring during a specific time and day because of speeding

Section 2: Line plot



Section 3: Descriptive statistics

Variable	OBS	Mean	Std Dev.	Min	Max
count	148	44.02	249.7	40	1199
Hour_im	148	11.9	7	0	23
Day_week	148	4	2	1	7
crash	148	.39	.5	0	1

Section 4: Logistic regression model for probability of fatal crashes because of speeding given time, day of the week, and frequency of fatal crashes occurring in the South region.

Table 3. Logit and Probit estimates	Model 1 Logit (MLE)	Model 2 Probit (MLE)
	Crash	Crash
*Hour_im (imputed hour of fatal accident)	-0.095 (0.07)	-0.05* (0.03)
*Day_week (the day that the fatal accident occurred)	-0.34 (0.26)	-0.13 (0.11)
Count	0.3*** (0.07)	0.123*** (0.02)
Constant	-11.13*** (3.03)	-4.6*** (1.05)
Log-likelihood value	175.19	171.2
Pseudo R-squared	.69	.68
% Correctly Predicted	99.3	99.3
AIC	174.6	170.8
Notes:	Selected logit model used to interpret the situation *indicates variable is not significant for Wald Chi-square value. *=p<0.1, ** = p<0.05, *** = p<0.01	