Statistical Analysis of Crash Factors Related to Auto Insurance

Paul E. Patterson, Matthew Kuofie

Grand Canyon University, USA

mr_ppatterson@hotmail.com,matthew.kuofie@my.gcu.edu, mhkuofie@msn.com

Abstract

Are the age of a person and their driving history key factors in the establishment of risk for auto insurance? The Center for Disease Control and Prevention (CDC) affirmed that American spend more than a million days in the hospital resulting from injuries sustained in motor vehicle accidents, $18 billion in medical costs, and productivity loss from work costing $33 billion. On average, 2.5 million Americans went to the emergency room, where 10% became hospitalized (2019). Statistical Analysis of Crash Factors attempts to determine if age and mileage are valid variables of risk in insurance underwriting to assess the level of coverage and cost. Risk is further validated using the mean, variance, and regression testing of the sample of groups. Key findings indicate the 80+ age group was more prone to die in a car accident; the 60-69 age group were less likely to die and is considered the safest age group surveyed. More startling, after the age of 69 risk doubles up 72% for the 70-79 age group to over 270% for persons 80 and over.

Keywords**:** auto insurance, risk factors, age, crashes, fatalities

A Statistical Analysis of Crash Factors Related to Auto Insurance

Introduction

According to the Center for Disease Control (CDC), six million car accidents resulting in two million injuries are reported annually (2019). Statistically, for every person in the United States, 327 million as of 2018 (The World Bank, n.d.), nearly four (3.8) people are more than likely to be involved in a traffic accident every minute, whether they are behind the wheel or not, and the probability of that accident being a fatality is close to four persons (3.65) every hour (NHTSA National Center for Statistics and Analysis, 2019) due to alcohol (40%), speed (30%), or recklessness (33%) (CDC. 2019).

Whether speed is a contributing factor, or other factors like drunk driving, distraction, or sleep deprivation, one point that remains car accidents are on the rise, especially in areas of high congestion (location) which is why many insurance agencies are adopting a pay-as-you-drive (PAYD), incentivizing loyal while maintaining market competition (Ayuso, Guillen, & Nielsen, 2016, p.4).

Purpose (Rationale)

The purpose of this paper was to bring to light, from a statistics standpoint, several factors affecting auto insurance – age of drivers, miles (driven), total crashes, and reported injuries concerning fatalities – probability, the likelihood that something will happen (Lind-Marchal-Wathen, *(2018)*. The hypothesis could then derive if age and mileage, or both are legitimate variables in determining risk, and overall exposure to higher insurance premiums.

Data Collection

Research and data were found using various available databases online from the AAA Foundation.org with supplemental information from Center for Disease Control and Prevention

(CDC), National Association of Insurance Commissioners (NAIC), the National Highway Traffic Safety Administration (NHTSA), and the U.S. Department of Health and Human Services National Library of Medicine. The raw data was then compiled using Minitab to generate a descriptive statistical summary report, charts, and graphs.

Factors

Insurance underwriting is the method used in pricing to assess risk leading to an assigned rating based on personal information and claims history, more specific, the severity of the accident's history (NAIC, 2019). These are the main possible factors that can affect the amount and price of insurance:

- Location
- Age of driver
- Gender
- Marital Status
- Driving Experience
- Driving Record
- Claims history
- Credit history
- Previous insurance coverage (under/uninsured)
- Vehicle age
- Vehicle type
- Vehicle use (work/recreational)
- Miles per 100M (driven to/from per day)
- Coverages and deductibles

The two main factors, Age of Driver and Miles Driven per million miles, will be used in this report to illustrate risk and its severity from automobile accidents.

Data Analysis and Results

There is a challenge in determining the actual algorithm used by insurance companies, perhaps for proprietary reasons, when rating an individual for insurance. One can, however, discern that from this list, data analysis is used to determine the mean, median, standard deviation, hence the Descriptive statistics utilized in the determination of price.

One method for determining price is to use the Frequency Model, which takes into consideration the frequency of occurrence and its severity. Following a Poisson distribution, the function of a linear function is as follows (Ayuso, Guillen, & Nielsen, 2016, p.5):

Table 1. Traffic Accident Crashes and Fatalities, AAAFoundation, 2018.

Age of Driver	All Crashes	Injury Crashes	Fatal Crashes
Rate per 100 million miles driven			
16-17	1,432	361	3.75
18-19	730	197	2.47
20-24	572	157	2.15
25-29	526	150	1.99
30-39	328	92	1.20
40-49	314	90	1.12
50-59	315	88	1.25
60-69	241	67	1.04
70-79	301	86	1.79
80+	432	131	3.85

Table 1. Traffic Accident Crashes and Fatalities, AAAFoundation, 2018.

Driver in age groups 16-17 continues to have the highest rate of involvement in crashes (1432 total crashes / 361 injury-related) resulting in injury to themselves and others, to include death of others in vehicular accidents in which they were involved. Though fatality-rate for this group is high, at 3.75, the 80+ age group suffers the highest driver-mortality at 3.85, reports the AAA Foundation for Traffic Safety (2018). The 16-17 and the 80+ age group are 1 ½ to three times more likely to die in a car crash any other age group. More surprising is that crashes, their injuries, and fatalities as a result of accidents drop significantly at the 30-39 age group. The safest group overall is the 60-69 age group with 1.04 fatalities, 361 crashes, and 67 injuries, which explains how insurance rates can adjust greatly from one age group to the next.

By measuring the variance and dispersion or the mean, median, and standard deviation of data, the goal was to better understand the accuracy of information (Minitab, n.d):

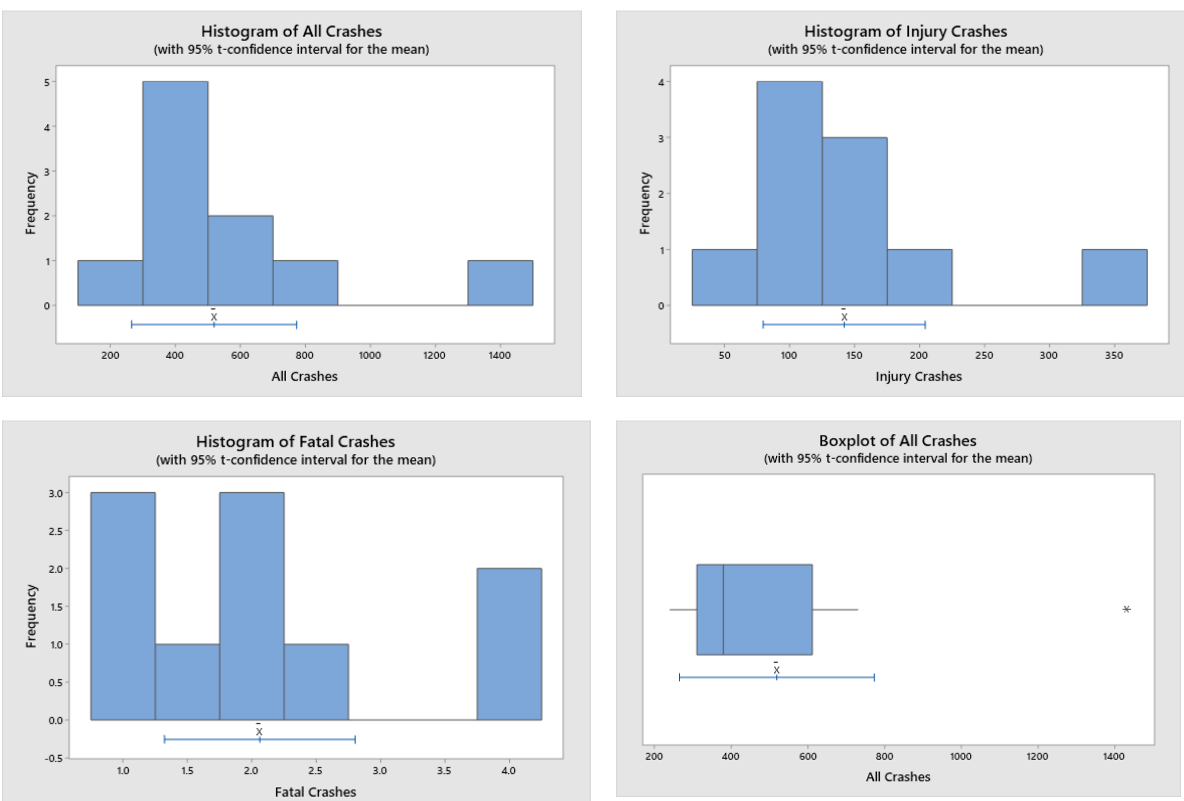
Table 2: One-Sample T: All Crashes, Injury Crashes, Fatal Crashes

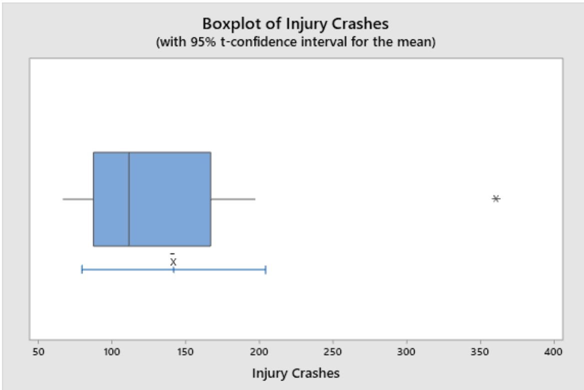
Descriptive Statistics

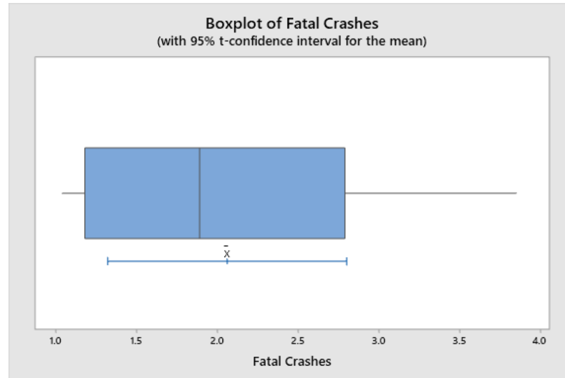
Sample	N	Mean	StDev	SE Mean	95% CI for μ
All Crashes	10	519	355	112	(265, 773)
Injury Crashes	10	141.9	87.1	27.5	(79.6, 204.2)
Fatal Crashes	10	2.061	1.035	0.327	(1.321, 2.801)

Note: μ: mean of All Crashes, Injury Crashes, Fatal Crashes

Figures 1: Histograms, Minitab, 2018.







A Mean of 519 indicates of the more than 5000 auto crashes the majority rests at around 500, but what is more interesting is the accident follows a bell curve peaking at 150 crashes. Fatalities, however, follow a more scattered approach with the highest level of fatalities occurring at the 1.0 and 2.0 marks. The boxplot graph shows the range is 400-600 with a mean of 519.

Here we tested the continuous and normal distributions for standard deviation (Minitab, n.d):

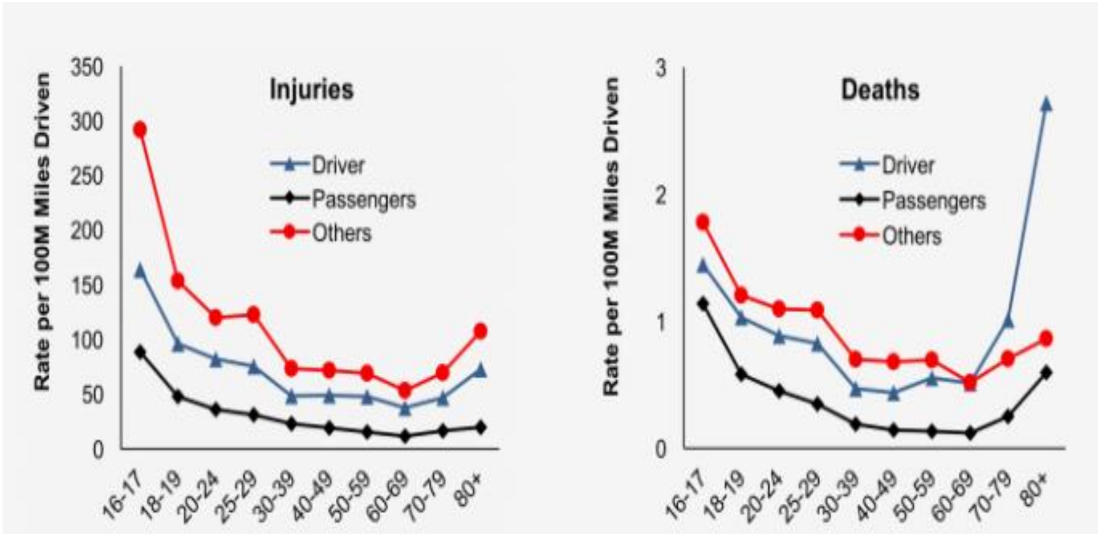
Table 3: Descriptive Statistics: Variance Output

Variable	N	StDev	Variance	95% CI for σ using Bonett	95% CI for σ using Chi-Square
All Crashes	10	355	125939	(128, 1227)	(244, 648)
Injury Crashes	10	87.1	7586	(33.4, 282.4)	(59.9, 159.0)
Fatal Crashes	10	1.03	1.07	(0.58, 2.28)	(0.71, 1.89)

Variance, which measures mean values and how they vary as a collection (Lind-Marchal-Wathen, 2018, p.71), is more accurate than mean. In this case, the Variance of all crashes is 125,939, 7586 are injury-related, more than 12k fatal.

When compared with the analysis from the AAA Foundation for Traffic Safety, the results were more revealing (2018). Figure 2 shows the number of driver-related deaths increases in the 70-79 age group.

Figure 2. Injuries versus Deaths per 100M Miles Driven, AAA Foundation.org, 2018.



In performing Regression Analysis, the equation for All Crashes is as follows (Minitab, n.d):

Table 4: Regression Analysis: All Crashes versus Injury Crashes

The regression equation is
All Crashes = - 57.93 + 4.066 Injury Crashes

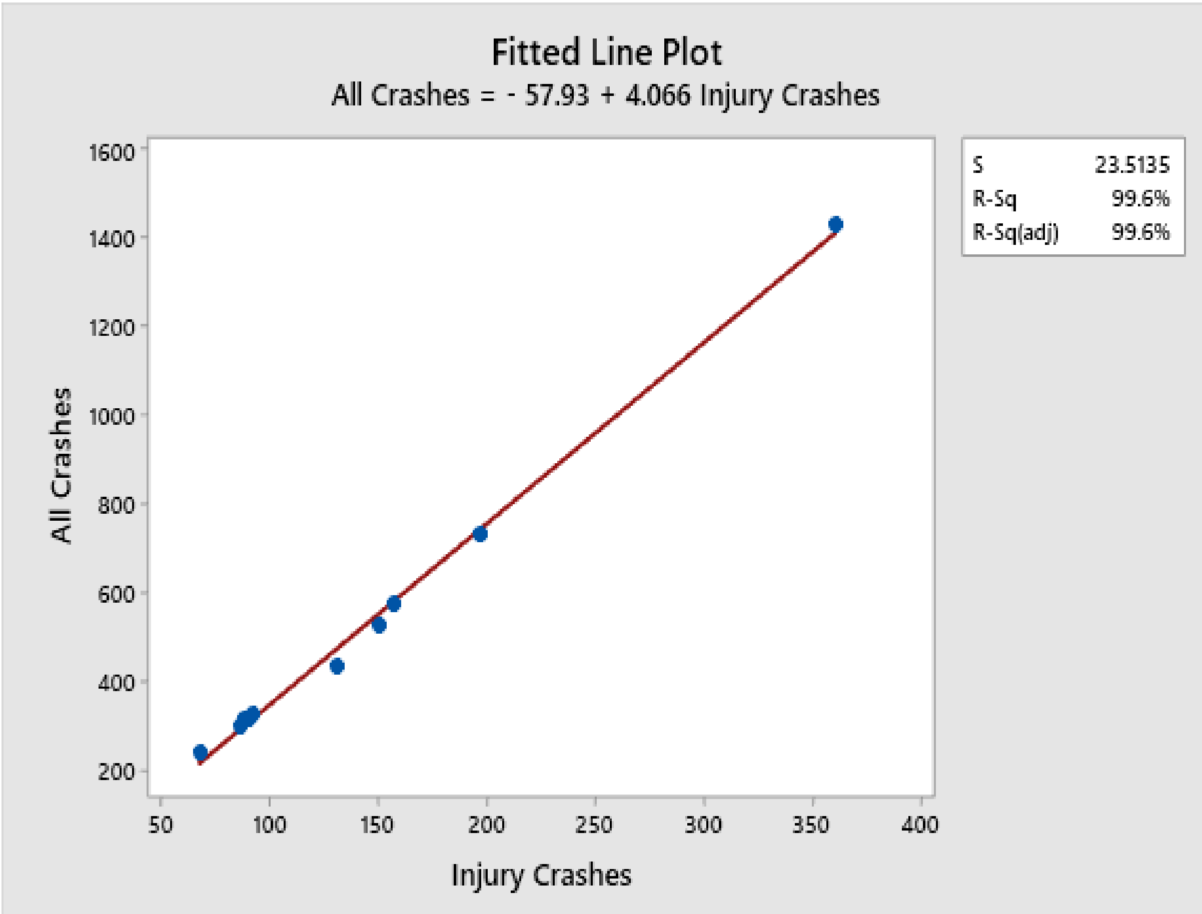
Model Summary

S	R-sq	R-sq(adj)
23.5135	99.61%	99.56%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1129024	1129024	2042.06	0.000
Error	8	4423	553		
Total	9	1133447			

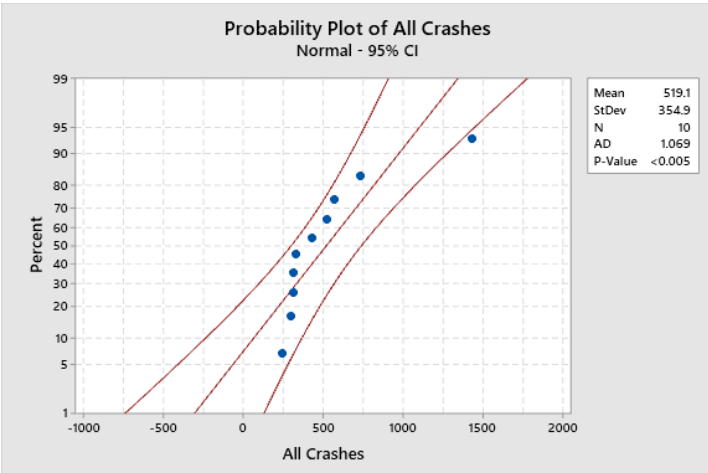
Figure 3: Regression Analysis of Injuries versus Deaths per 100M Miles Driven, AAA Foundation.org, 2018.

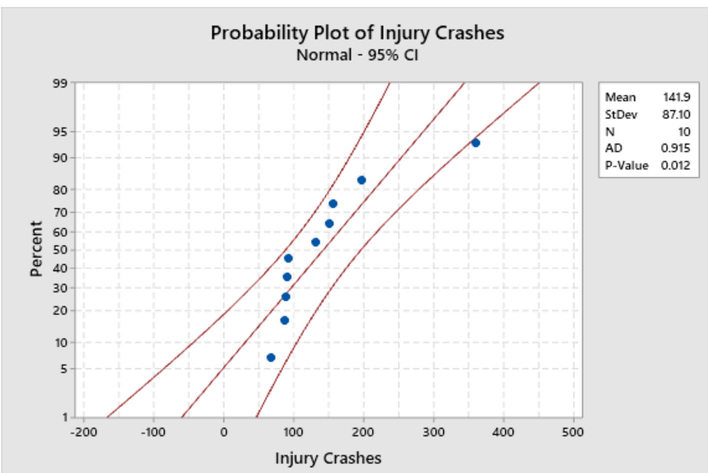


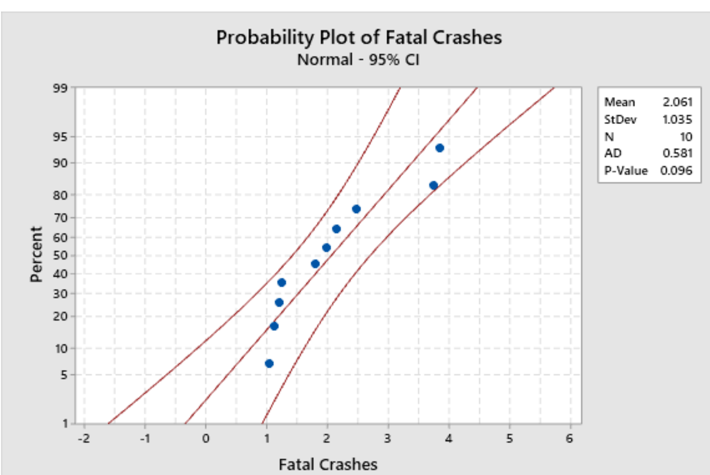
The fitted Line Plot, regression analysis, displays the relationship of a predictor and its response (Minitab.com, n.d.). All Crashes is the predictor while the responses are Injury Crashes and Fatality Crashes (not shown) (Minitab, n.d.):

Probability Plot of All Crashes, Injury Crashes, Fatal Crashes

Figures 4-6: Regression Analysis of Injuries versus Deaths per 100M Miles Driven, AAA Foundation.org, 2018.







Conclusion

Certain risk factors, such as the age of driver and miles driven, play a significant role in understanding risk and the likelihood, and survivability, of a person involved in a serious car accident. Risk equates to probability of occurrence which quantifies how much insurance a person needs, or pays. The null hypothesis used was whether any age group were more prone than the other, specifically the 16-17 age groups.

Conclusively, the mean and variance with a 95% confidence level indicated that the 80+ age group is more likely to die from a car accident while the safest are people near the age of retirement (60-69), which seem to be the only likely correlation since key factors indicates risk doubling every 10 years from 1.79 fatalities for 70-79 to 3.85 for 80+. Therefore, insurance companies, having the means, the data, and knowledge, can determine the level of insurance and its price for one age group to the next.

References

AAAFoundation. (2018, June 14). Rates of motor vehicle crashes, injuries, and deaths in relation to driver age, United States, 2014-2015. Retrieved from https://aaafoundation.org/rates-motor-vehicle-crashes-injuries-deaths-relation-driver-age-united-states-2014-2015/

Ayuso, M., Guillen, M., & Nielsen, J. P. (2016, August). Improving automobile insurance ratemaking using telemantics: Incorporating mileage and driver behaviour data. Retrieved from https://www.recercat.cat/bitstream/handle/2072/270305/XREAP2016-08.pdf?sequence=1

Center for Disease Control and Prevention (CDC). (2019, February 12). Motor vehicle crash injuries. Retrieved from https://www.cdc.gov/vitalsigns/crash-injuries/index.html

Lind-Marchal-Wathen, *(2018). Statistical techniques in business and economics* (17th ed.). McGraw-Hill

Minitab.com. (n.d.). Help and how-to overview. Retrieved from https://support.minitab.com/en-us/minitab/18/help-and-how-to/help-and-how-to-overview/NHTSA National Center for Statistics and Analysis. (2019, April). Summary of statistical findings. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812700

National Association of Insurance Commissioners (NAIC) | Center for Insurance Policy and Research. (2019, May). Auto insurance. Retrieved from https://www.naic.org/cipr_topics/topic_auto_insurance.htm

The World Bank. (n.d.). Population, total. Retrieved from https://data.worldbank.org/indicator/SP.POP.TOTL