1 Abstract

Due to the alarming number of traffic related deaths in 2018, a statically analysis is performed in order to identify any possible variables associated with traffic fatalities. Using a dataset named “Fatalities” with information gathered from multiple US agencies, a multiple regression model was fitted using selection methods to determine a model for the rate of traffic fatalities. From this model, it can be shown that several key variables are strongly associated with the rate of traffic fatalities. From this information, politician and other policy makers can use these trends and associations as informative tools to help create and reform legislation to better reduce traffic related fatalities.

2 Introduction

According to the National Highway Traffic Safety Administration, it was reported that in 2018 there were 36,560 highway fatalities in 2018 which sparked an interest into analyzing what policies or characteristics may be leading to these fatalities (NHTSA, 2019). Aside from the 36,000 traffic fatalities in 2018 alone, the CDC found that traffic fatalities are one of the leading causes of death in the United States for people between the ages of 1 and 54 (CDC, 2020).

In order understand these relationships, a dataset made up of the culmination of information from various US agencies is implemented to detect if there is an association between any of the variables provided and vehicle fatalities. This dataset contains many variables including drinking age, beer tax rates, local unemployment rates as well as binary variables such as if there is mandatory jail time or community service consequences in place for drivers caught driving drunk or recklessly. From these variables, a new rate variable for the number of car fatalities over a given population is made and used as the response variable in a multiple regression model to determine if any of the remaining variables have strong associations with the rate of traffic fatalities.

From this analysis and its conclusions, policymakers can find statistical evidence and be influenced by the identified associations between the rate of traffic fatalities and the predictor variables of the model to make changes to current traffic regulations that lessen and prevent traffic related deaths. As such, the proposed hypothesis of this analysis is that a multiple regression model can be found that accurately models the rate of traffic fatalities with the remaining variables that shows strong associations between certain variables and the rate of traffic fatalities.

3 Background

The dataset used in this analysis is cross-sectional time series data sourced from several different US agencies including the US Department of Transportation, the US Bureau of Economic Analysis, and the US Bureau of Labor Statistics among others and included data that varied from local unemployment rates and traffic violation consequences to the overall percentage of Mormons in a given state (Kleiber, 2008). The data contains 336 observations over 34 variables that were acquired over several years spanning from 1982 to 1988 from every state in the United States besides Hawaii and Alaska (Kleiber, 2008). Since the data comes from a stratified sampling method where states and years are the stratas, it’s adequate to deduce that the target population of this dataset is the entire United States.

Important variables used in the analysis include a new variable that was formulated from existing variables, “rate,” which is used as the response variable in the model and represents the rate of traffic fatalities for a given population. Other important variables used for the predictors of the model include “income” which represents per capita personal income in 1987 dollar values, “miles” which is the average amount of miles for drivers, “service” which is binary and refers to if community service is a mandatory sentence for unsafe drivers, and “Baptist” or “Mormon” which refer to the overall percentage of Baptists or Mormons in a given population. There are a few other variables used in the modelling process but they are relatively self-explanatory.

References

Center for Disease Control and Prevention (2020). Road Traffic Injuries and Deaths—A Global Problem. Retrieved February 11, 2021, from https://www.cdc.gov/injury/features/global-road-safety/index.html

Kleiber C, Zeileis A (2008). Applied Econometrics with R. Springer-Verlag, New York. ISBN 978-0-387-77316-2, <https://CRAN.R-project.org/package=AER>.

NHTSA (2019). U.S. Transportation Secretary Elaine L. Chao Announces Further Decreases in Roadway Fatalities [Editorial]. NHTSA.gov. Retrieved February 11, 2021, from <https://www.nhtsa.gov/press-releases/roadway-fatalities-2018-fars>