Final Project Car Accidents

Chad Stephens

5/6/2021

## Introduction

This assignment is the final project for the IS 4300 R coding class. Each student chose a data base to study and the provide analyisi to Dr. Josh Holland. For this project the dataset we will be using is the DUI dataset. The DUI dataset is one that is used to track the number of DUI’s per state and fatalities by car accident per state. This dataset has six variables that are shown below. Those variables are State, DUI(Driving Under the Influence),Fatalities, Population, Sunday.Sales, and fat\_rate. The ones we will be analyzing is **fat\_rate** which is the fatality rate per 10000 people, **Population**, and **DUI**(Driving Under the Influence charges). We will be trying to find if there is any correlation between fatality rate and other variables in this dataset.

dui <- read.csv(file = "C:/Users/Chad Stephens/Desktop/Excel Data Workbooks/DUI.csv",header = TRUE)  
DUI\_p <- as.data.frame(dui)  
#this makes the dataset into a data frame  
DUI\_p <- DUI\_p %>% mutate(fat\_rate = Fatalities/10000) #fat\_rate is fatalities 10000 people  
head(DUI\_p)

## State DUI Fatalities Population Sunday.Sales fat\_rate  
## 1 Alabama 7863 247 4858979 Prohibited 0.0247  
## 2 Alaska 3163 24 738432 Permitted 0.0024  
## 3 Arizona 22367 272 6828065 Permitted 0.0272  
## 4 Arkansas 6919 149 2978204 Local 0.0149  
## 5 California 141458 914 39144818 Permitted 0.0914  
## 6 Colorado 25562 151 5456574 Permitted 0.0151

## Analysis

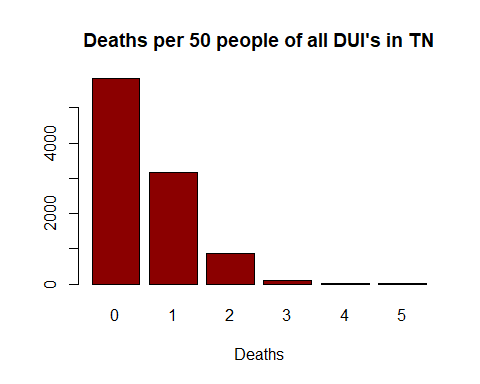
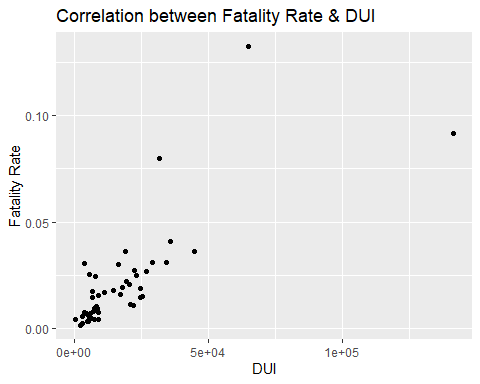
This is the a two sample t-test to find if there is a significant level of correlation between the amount of DUI’s per state and the fatality rate per state. As shown below there is a significant relationship between the two variables because the p-value is well **below** .05 at 1.003e-06. Because the t-test has a p-value below .05 this means the null hypothesis should be rejected. This analysis claims that as DUI’s increase fatality rate increases. This shows that these two variables go hand in hand.

t.test(DUI\_p$DUI,DUI\_p$fat\_rate)

##   
## Welch Two Sample t-test  
##   
## data: DUI\_p$DUI and DUI\_p$fat\_rate  
## t = 5.5878, df = 49, p-value = 1.003e-06  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 11089.41 23545.22  
## sample estimates:  
## mean of x mean of y   
## 17317.340000 0.020538

## Visualization

This visualization shows the correlation between DUI and the fatality rate. As shown in the graph there are a few outliers with higher fatality in comparison to the number of DUI’s. As shown by this graph it shows the correlation that as DUI’s increase the fatality rate seems to increase. The second is a visualization of a sample of deaths by DUI in TN. This barplot shows the probability of at least one death per 50 DUI’s in the state of TN. Although 0 still has the highest probabilty the likely hood of a death is still very high.



## Conclusion

To conclude using the fatality rate which is deaths per every 10000 people and DUI’s there is a correlation that shows as DUI’s increase the fatality rate increases. It is important to not that correlation does not always mean causation, but it is note worthy how low the p-value was for this t-test.