Final Report

Tong Wu

4/30/2021

```r
modeldata <- read.csv("modeldata.csv")
```

```r
modeldata <- modeldata %>%
  filter(party_age %in% c(16:100)) %>%
  mutate(party_race = as.factor(party_race)) %>%
  mutate(vehicle_type_sort = as.factor(vehicle_type_sort))
modeldata$party_race <- relevel(modeldata$party_race, ref="white")
modeldata$vehicle_type_sort <- relevel(modeldata$vehicle_type_sort, ref="other")
```

```r
modeldata <- modeldata %>%
  filter(carnation != "OTHER")
```

```r
modeldata <- modeldata %>%
  mutate(alcohol_involvedstatus = case_when(
    alcohol_involved == 0 ~ "no",
    alcohol_involved == 1 ~ "yes"
  ))
```

1. Introduction

1.1 Background

Alcohol-related car accidents are a major problem in the United States. In 2016, 10,497 people died of alcohol-related car accidents and accounted for 28% of all traffic-related deaths. Extensive studies have investigated how the odds of alcohol-related vehicular crashes are associated with factors such as race, age, car types, and socioeconomic status. Younger drivers are associated with an increase in risk of alcohol-related accidents in both sexes [1]. Alcohol-related accidents are also more common among male drivers than female drivers. [2] Latino drivers have consistently had a higher prevalence of being involved in drunk driving crashes. RAM 2500, Chevy S-10, and BMW 4-series were found to have the most DUIs according to data from Insurance company Insurify. [3] Furthermore, lower socioeconomic status was associated with increased risks of drunk driving. [4] However, there hasn't been a definitive study that, to our knowledge, accounted for all of these factors, along with other factors such as weather and road conditions, specifically in alcohol-impaired motor crashes.

The aim of our study is to determine significant factors that are associated with an increase in the probability of alcohol-related car accidents in California through constructing a statistical model. We are interested in the following questions:

1. Are certain car characteristics, such as the model year, car type, nationality of the car brand, whether the car is a premium brand or classified as a sports car, associated with increased odds of alcohol-related accidents?

2. Are demographic factors such as race, sex, and age significantly associated with the probability of alcohol-related accidents?

3. Is insurance status associated with lower risks of drunk-driving accidents, and how does this association vary by race, sex, age, and whether a driver has a premium car or a sports car?

We want to examine these factors while also controlling for weather conditions, crash types, and county population. We hope that our analysis will provide helpful insights for insurance companies and policymakers to develop solutions to combat alcohol-related car accidents.

1.2 Data Information

The dataset we used was from the California Traffic Collision Data from Statewide Integrated Traffic Records System (SWITRS). We only focused on automobile accidents that took place in 2020, which totaled 2,741,357 events. The dataset contains various information about each accident, including characteristics of the driver, insurance status, car involved, whether alcohol was involved, county in which the crash took place, as well as the weather and road conditions during the crash. We decided to focus only on crash events where the driver was at fault, because we were interested in understanding what factors about the driver are associated with increased rates of alcohol-related car accidents. Since all the observations only occurred in the state of California, the insights we gain from our analysis will not be generalizable to areas of the US with very different traffic laws, weather conditions, and economy/weath.

1.3 Exploratory Data Analysis

Our response, or dependent variable, is a binary variable of whether the accident was alcohol-related or not.

To get a better understanding of vehicle characteristics associated with drunk driving crashes for the first question, we examined the vehicle's model year, type, nationality of the vehicle brand, whether the vehicle was premium, and whether the vehicle was classified as a sports car. The nationality of the vehicle brand was grouped into American, Asian, and European. Cars classified as premium included brands such as Mercedes, Lexus, and Audi. Cars classified as sports cars included brands such as Ferrari, Lamborghini, and Aston Martin. Vehicle type was simplified from the original dataset by classifying vehicles as only passenger cars, trucks, two-wheeled vehicles such as motorcycles, buses, and other vehicle types.

To examine demographic characteristics about the driver for the second question, we examined the age, race, and sex of the drivers at fault in the crash and determined whether certain characteristics are significantly associated with higher rates of alcohol-related accidents.

For the third question, we examined whether the driver is insured or not, and determined whether it is significantly associated with drunk crashes. In addition, we are interested in determining if significant two-interactions exist between insurance status and race, sex, age, and premium car status because insurance coverage rates could vary by these factors.

Finally, we also controlled for environmental factors, such as weather and road surface conditions, that can affect general car crash rates in our model. In addition, we also controlled for county population because it is reasonable to assume that higher density counties will have more accidents and thus more alcohol-related vehicle crashes.

1.3 Exploratory Data Analysis

To visualize how the rate of alcohol related accidents varied across the variables of interest, we generated stacked bar plots. Below are plots:

```
model4 <- glm(data=modeldata, alcohol_involved ~ party_sex * financial_responsibility + party_age * fina
```

```r
summary(model4)
```

```
##
## Call:
## glm(formula = alcohol_involved ~ party_sex * financial_responsibility +
##     party_age * financial_responsibility + party_race * financial_responsibility +
##     vehicle_year + vehicle_type_sort + premiumcar * financial_responsibility +
##     carnation + sportcar * financial_responsibility + population +
##     weather_1 + type_of_collision, family = binomial(logit),
##     data = modeldata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5256  -0.4883  -0.3755  -0.2879   3.1001
##
## Coefficients:
##                                              Estimate Std. Error z value
## (Intercept)                                 -2.632e+00  5.552e-01  -4.740
## party_sexmale                                5.782e-01  1.001e-02  57.763
## financial_responsibilityY                   -7.044e-01  1.736e-02 -40.586
## party_age                                    4.525e-03  3.203e-04  14.126
## party_raceasian                             -2.856e-01  2.775e-02 -10.290
## party_raceblack                             -3.877e-01  1.472e-02 -26.328
## party_racehispanic                           1.597e-01  9.926e-03  16.087
## party_raceother                             -2.972e-01  2.356e-02 -12.612
## vehicle_year                                -5.499e-05  2.741e-04  -0.201
## vehicle_type_sortbus                        -1.501e-01  1.058e-01  -1.419
## vehicle_type_sortpassenger car               5.927e-01  9.311e-02   6.366
## vehicle_type_sorttruck                       4.632e-01  9.317e-02   4.971
## vehicle_type_sorttwo wheel                  -2.531e-01  9.561e-02  -2.647
## premiumcaryes                               -1.975e-02  1.299e-02  -1.520
## carnationASIAN                               6.958e-03  4.490e-03   1.550
## carnationEUROPEAN                            9.376e-02  8.902e-03  10.532
## sportcaryes                                  3.827e-01  1.250e-01   3.063
## population                                  -1.407e-02  1.065e-03 -13.213
## weather_1cloudy                             -2.942e-01  6.312e-03 -46.611
## weather_1fog                                 7.551e-02  2.835e-02   2.664
## weather_1other                              -2.325e-01  7.853e-02  -2.961
## weather_1raining                            -2.881e-01  1.098e-02 -26.251
## weather_1snowing                            -1.260e+00  7.108e-02 -17.724
## weather_1wind                                8.687e-02  1.121e-01   0.775
## type_of_collisionhead-on                     1.062e+00  9.836e-03 107.939
## type_of_collisionhit object                  1.544e+00  6.951e-03 222.183
## type_of_collisionother                       4.999e-01  1.876e-02  26.647
## type_of_collisionoverturned                  1.222e+00  1.182e-02 103.448
## type_of_collisionpedestrian                  2.705e-01  1.934e-02  13.983
## type_of_collisionrear end                    2.527e-01  6.920e-03  36.522
## type_of_collisionsideswipe                   7.647e-01  7.683e-03  99.539
## party_sexmale:financial_responsibilityY      7.164e-03  1.118e-02   0.641
## financial_responsibilityY:party_age         -1.330e-02  3.498e-04 -38.016
## financial_responsibilityY:party_raceasian   -3.457e-01  2.976e-02 -11.619
## financial_responsibilityY:party_raceblack    2.607e-01  1.741e-02  14.977
## financial_responsibilityY:party_racehispanic 3.000e-03  1.112e-02   0.270
## financial_responsibilityY:party_raceother   -3.174e-01  2.612e-02 -12.151
```

```
## financial_responsibilityY:premiumcaryes      1.347e-01  1.406e-02   9.575
## financial_responsibilityY:sportcaryes       -2.090e-01  1.338e-01  -1.562
##                                              Pr(>|z|)
## (Intercept)                                  2.14e-06 ***
## party_sexmale                                < 2e-16 ***
## financial_responsibilityY                    < 2e-16 ***
## party_age                                    < 2e-16 ***
## party_raceasian                              < 2e-16 ***
## party_raceblack                              < 2e-16 ***
## party_racehispanic                           < 2e-16 ***
## party_raceother                              < 2e-16 ***
## vehicle_year                                 0.84097
## vehicle_type_sortbus                         0.15592
## vehicle_type_sortpassenger car               1.94e-10 ***
## vehicle_type_sorttruck                       6.67e-07 ***
## vehicle_type_sorttwo wheel                   0.00811 **
## premiumcaryes                                0.12852
## carnationASIAN                               0.12126
## carnationEUROPEAN                            < 2e-16 ***
## sportcaryes                                  0.00219 **
## population                                   < 2e-16 ***
## weather_1cloudy                              < 2e-16 ***
## weather_1fog                                 0.00773 **
## weather_1other                               0.00307 **
## weather_1raining                             < 2e-16 ***
## weather_1snowing                             < 2e-16 ***
## weather_1wind                                0.43823
## type_of_collisionhead-on                     < 2e-16 ***
## type_of_collisionhit object                  < 2e-16 ***
## type_of_collisionother                       < 2e-16 ***
## type_of_collisionoverturned                  < 2e-16 ***
## type_of_collisionpedestrian                  < 2e-16 ***
## type_of_collisionrear end                    < 2e-16 ***
## type_of_collisionsideswipe                   < 2e-16 ***
## party_sexmale:financial_responsibilityY      0.52154
## financial_responsibilityY:party_age          < 2e-16 ***
## financial_responsibilityY:party_raceasian    < 2e-16 ***
## financial_responsibilityY:party_raceblack    < 2e-16 ***
## financial_responsibilityY:party_racehispanic 0.78727
## financial_responsibilityY:party_raceother    < 2e-16 ***
## financial_responsibilityY:premiumcaryes      < 2e-16 ***
## financial_responsibilityY:sportcaryes        0.11840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1931434  on 2741356  degrees of freedom
## Residual deviance: 1729840  on 2741318  degrees of freedom
## AIC: 1729918
##
## Number of Fisher Scoring iterations: 5
```