CAR CRASH ANALYSIS

**Introduction**

Over the last decade and more the automotive safety standards have considerably gone up. With introduction of collision control and lane departure alert systems, the range of electronic safety systems have added the much-required safety features to ensure less fatal accidents. Considering that, We’ve analyzed 2016 – 2018 crashes in US.

Aim

The project will analyze various contributing factors like drowsiness, distracted driver, use of alcohol, use of restraint (seat belts) etc that lead to fatalities. The aim of that analysis would be to determine factors contributing to the increase in the probability of a crash turning fatal.

Data

This data is provided by NHTSA (National Highway Traffic Safety Administration). NHTSA uses data from many sources, including Crash Report Sampling System (CRSS). The CRSS obtains its data from a nationally representative probability sample selected from more than six million police reported crashes which occur annually. The data contains crash data of years 2016 -2018 with information about Accident, Person, Vehicle. This data allows our team to analyze following fields on Severity of Injury, Fatal, Day of week, hour of accident, interstate highway, light condition, alcohol abuse, weather, deformed car, sex, type of intersection etc.,

Data Url: <https://www.nhtsa.gov/node/97996/221>

**Exploratory data analysis**

**Inference**

Null Hypothesis : there’s no impact on fatality of injury based on accident attributes such as light condition, alcohol, interstate highway, hour of accident , day of week, deformed car, sex of driver, type of intersection.

Alternate Hypothesis: there is a significant impact of accident attributes on fatality of injury

> anova(model, test= "Chisq")

Analysis of Deviance Table

Model: binomial, link: logit

Response: fatalornonfatal

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 99239 18811

INT\_HWY 1 35.87 99238 18775 2.115e-09 \*\*\*

ALCOHOL 1 347.84 99237 18427 < 2.2e-16 \*\*\*

WEATHER 1 2.92 99236 18424 0.08769 .

TYP\_INT 1 126.19 99235 18298 < 2.2e-16 \*\*\*

LGT\_COND 1 91.03 99234 18207 < 2.2e-16 \*\*\*

SEX 1 275.11 99233 17932 < 2.2e-16 \*\*\*

DEFORMED 1 201.38 99232 17730 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Best Model**

glm(fatalornonfatal~ INT\_HWY + ALCOHOL + WEATHER + TYP\_INT +LGT\_COND + SEX + DEFORMED, data = allYearAcc\_train, family=binomial)

**Prediction**

**Conclusion**

References:

1. Preusser, D. F., Williams, A. F., & Ulmer, R. G. (2000, January 27). Analysis of fatal motorcycle crashes: crash typing. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/0001457595000275>
2. Yasmin, S., Eluru, N., & Pinjari, A. R. (2015, November). Pooling data from fatality analysis reporting system (FARS) and generalized estimates system (GES) to explore the continuum of injury severity spectrum. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/26342892>

**Team:**

Karmanya Pathak

Sharath Chandra Raini

Rebecca Adedeji

Anand Saurabh

Mrugank Salunke