Recipe for A Disaster Behind the Wheel: Predicting Car Crash Fatality Rates

[redacted]

INTRODUCTION

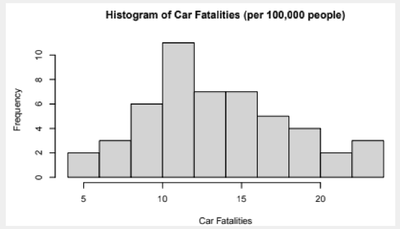
- Car accidents are a significant concern in the United States; leading cause of death in the US, with an average of 35,791 deaths each year from 2015-19 (CDC, 2022)
- Possible external factors related to the fatality of car crashes led our team to consider characteristics such as weather conditions and legal speed limits that could pose dangers (DOT, 2024)
- Studying the factors that potentially contribute to higher fatality rates can help society better understand and possibly prevent the loss of lives on the road

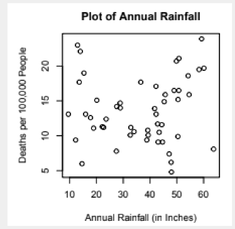
RESEARCH QUESTIONS

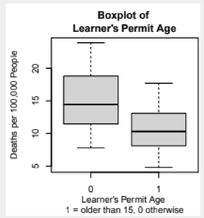
What Factors Affect Car Crash Fatality Rates?

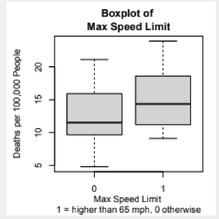
- On average, do states with a higher proportion of drivers in the 18 and 24 age group have a higher fatality rate (number of fatalities per 100,000 people) from car accidents?
- Do states with a lower maximum speed limit of 55 mph tend to have a lower fatality rate (number of fatalities per 100,000 people) from car accidents compared to the median maximum speed limit of all US states (65 mph)?
- Do states with higher average annual rainfall compared to the average among all states have a higher fatality rate in car accidents (number of fatalities per 100,000 people)?

EDA

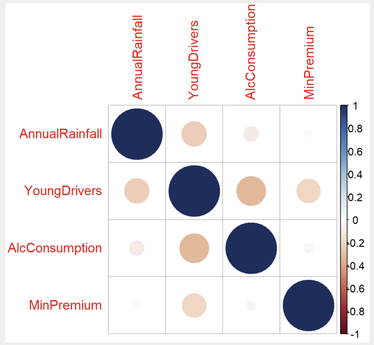








MULTICOLLINEARITY



- No severe multicollinearity
- Because all individual VIF values are below 10 and the average is less than 3, we deemed stepwise regression as unnecessary

```
AlcConsumption          MinPremium
         1.176               1.068

AnnualRainfall       YoungDrivers
         1.101               1.300
[1] 1.16125
```

MODEL BUILDING

Stage 1: Add Quantitative Variables, Higher Order Terms, and Quantitative-Quantitative Interactions

Initial Model: Fatality Rate $= \beta_0 + \beta_1 AnnualRainfall + \beta_2 AlcConsumption + \beta_3 MinPremium + \beta_4 YoungDrivers + \beta_5 YoungDrivers * MinPremium + \beta_6 AnnualRainfall^2$

Final Model: Fatality Rate $= \beta_0 + \beta_1 AnnualRainfall + \beta_2 AlcConsumption + \beta_3 MinPremium + \beta_4 AnnualRainfall^2$

Stage 2: Add Quantitative Variables

Initial Model: Fatality Rate $= \beta_0 + \beta_1 AnnualRainfall + \beta_2 AlcConsumption + \beta_3 MinPremium + \beta_4 AnnualRainfall^2 + \beta_5 Female\_Dummy + \beta_6 Permit\_Dummy + \beta_7 Mountain\_Dummy + \beta_8 Speed\_Dummy$

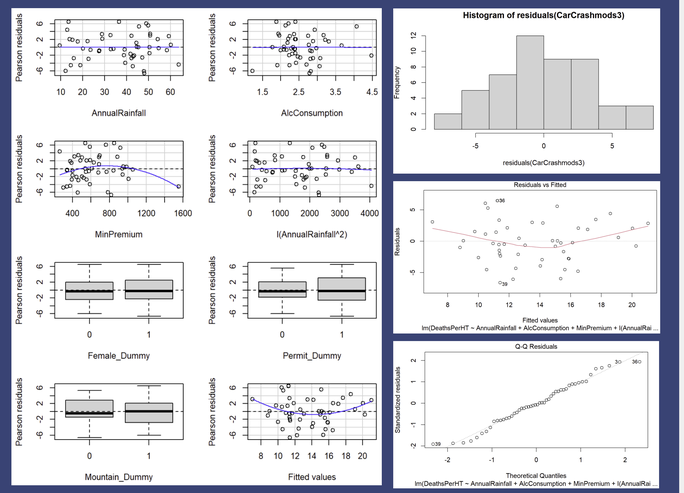
Final Model: Fatality Rate $= \beta_0 + \beta_1 AnnualRainfall + \beta_2 AlcConsumption + \beta_3 MinPremium + \beta_4 AnnualRainfall^2 + \beta_5 Female\_Dummy + \beta_6 Permit\_Dummy + \beta_7 Mountain\_Dummy$

Stage 3: Add Quantitative-Qualitative Interactions

Initial Model: Fatality Rate $= \beta_0 + \beta_1 AnnualRainfall + \beta_2 AlcConsumption + \beta_3 MinPremium + \beta_4 AnnualRainfall^2 + \beta_5 Female\_Dummy + \beta_6 Permit\_Dummy + \beta_7 Mountain\_Dummy + \beta_8 AnnualRainfall * Mountain\_Dummy$

Final Model: Fatality Rate $= \beta_0 + \beta_1 AnnualRainfall + \beta_2 AlcConsumption + \beta_3 MinPremium + \beta_4 AnnualRainfall^2 + \beta_5 Female\_Dummy + \beta_6 Permit\_Dummy + \beta_7 Mountain\_Dummy$

MLR ASSUMPTIONS



CONCLUSION

Final Model (Prediction Equation)

$$\widehat{Fatality\ Rate} = 28.777 - 0.952 AnnualRainfall + 0.022 AlcConsumption - 0.004 MinPremium + 0.014 AnnualRainfall^2 + 3.091 Female\_Dummy - 2.882 Permit\_Dummy + 0.793 Mountain\_Dummy$$

Interpretation:

- The positive quadratic term for Annual Rainfall highlights that both very low and very high rainfall levels are linked to higher fatality rates.
- An increase of 1 gal. ethanol consumption per capita increases the mean fatality rate by 0.174891.
- States with minimum permit age >15 have a mean fatality rate that is 3.863497 lower than the states with permit age <15.
- The mean fatality rate of states dominated by mountains (elevation range >3000) is 0.215877 higher than other states.

Efficacy:

- The model explains 48.74% of the variation in fatality rates.
- The adjusted R squared has improved (from 0.3332 to 0.3977) after removing observations 11 and 32.
- We used the PRESS statistic which indicated that the model's predictive accuracy could be improved.

Example:

- For a state with 30 inches of rainfall annually, 2 gallons of ethanol per capita, and all other variables equal to 0, the predicted fatality rate is approximately 12.86 deaths per 100,000 people.

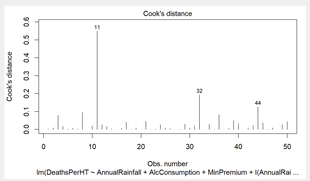
FUTURE RESEARCH/IMPROVEMENTS

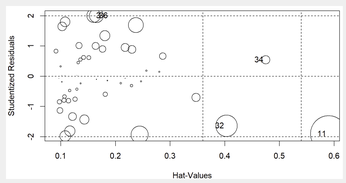
- Use larger datasets to extend model application to a broader country level.
- Add more variables to account for additional important factors like distracted driving and road quality.
- Explore time series data to account for time dependent factors and trends.
- Transform linear models to non-linear to achieve better accuracy.

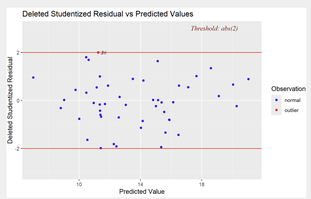
DATA SUMMARY

Abbreviated Name	Description	Levels/Units
DeathsPerHT (response)	The number of deaths (fatalities) per 100,000 people from car accidents in a state.	Deaths per 100,000 people
FemalePop	The proportion of female population in a state	N/A
Female_Dummy	Indicator if the state has a higher proportion of females.	If FemalePop >50% Female_Dummy = 1, otherwise Female_Dummy = 0
MaxSpeedLimit	The maximum posted speed limits for ordinary vehicles on urban interstates.	In miles per hour (mph)
Speed_Dummy	Indicator for high maximum speed limit	If MaxSpeedLimit >65 (median) Speed_Dummy = 1, otherwise Speed_Dummy = 0
PermitAge	The minimum legal age to obtain a permit.	In years
Permit_Dummy	Indicator for higher minimum permit age	If PermitAge >15 Permit_Dummy =1, otherwise Permit_Dummy = 0
ElevationRange	The range of elevation within a state, measured by taking the difference between the lowest and highest elevation in a state.	In feet
Mountain_Dummy	Indicator if the state is dominated by mountains	If ElevationRange >3000 Mountain_Dummy = 1, otherwise Mountain_Dummy = 0
AnnualRainfall	The average annual rainfall in the state	In inches
YoungDrivers	The proportion of drivers between the ages of 18 and 24	N/A
AlcConsumption	The alcohol consumption in a state	In gallons of ethanol per capita
MinPremium	The average annual cost of minimum car insurance coverage	In dollars

RESIDUALS







Cook's Distance Plot: 11, 32, and 44 are influential

Leverage Plot: 11, 32, and 34 are outliers in the x-direction

Deleted Studentized Residual Plot: 36 is an outlier

Observations 11 and 32 were removed to account for the violation of lack of fit

ADDITIONAL TECHNIQUE: PRESS

```
[1] 490.4464
[1] 721.6392
```

<- RSS

<- PRESS Statistic

FINAL MODEL & PREDICTION EQUATION

Adjusted R : 0.3977

P-Value: 0.0001955

RMSE: 3.502

$$\widehat{Fatality\ Rate} = 28.777 - 0.952 AnnualRainfall + 0.022 AlcConsumption - 0.004 MinPremium + 0.014 AnnualRainfall^2 + 3.091 Female\_Dummy - 2.882 Permit\_Dummy + 0.793 Mountain\_Dummy$$

WORKS CITED

1. Centers for Disease Control and Prevention. (2024, October 29). About transportation safety. https://www.cdc.gov/transportation-safety/about/index.html
2. U.S. Department of Transportation. (2024, June). Overview of Motor Vehicle Traffic Crashes in 2022. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813560
3. Denise, W. (2024, August 28). Driving Age by State. Parents. https://www.parents.com/driving-age-by-state-8607683
4. Educational Resources. Highest and Lowest Elevations. U.S. Geological Survey. https://www.usgs.gov/educational-resources/highest-and-lowest-elevations
5. Federal Highway Administration. (2024, March 5). Highway Statistics 2022. Office of Highway Policy Information. https://www.fhwa.dot.gov/policyinformation/statistics/2022/
6. Insurance Institute for Highway Safety. (2024, October). Maximum posted speed limits. https://www.iihs.org/topics/speed/speed-limit-laws
7. Megan, S. (2024, May 2). Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977-2022. National Institute on Alcohol Abuse and Alcoholism. https://www.niaaa.nih.gov/publications/surveillance-reports
8. Natalie, T. (2024, October 1). Car insurance rates by state for October 2024. Bankrate. https://www.bankrate.com/insurance/car/states/#minimum-car-insurance-in-each-state
9. Social Explorer. ACS 2022 (5-Year Estimates). https://www.socialexplorer.com/tables/ACS2022_5yr/R13745665
10. World Population Review. (2024). Wettest States 2024. https://worldpopulationreview.com/state-rankings/wettest-states
11. Yellman, M. A. (2022, July 1). Motor vehicle crash deaths - United States and 28 other high-income countries, 2015 and 2019. Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/71/wr/mm7126a1.htm