Predicting Violent Crime Rates

The 4th Side

Introduction

This semester, our team has decided to conduct a research project to find any variables that seem to have a relationship with crime rates in the United States. According to Statista, the US has seen an unmatched violent crime rate over the past couple decades, specifically with respect to homicides, when compared with rates in other countries. The United States has the highest crime rate of high-income countries by a wide margin. According to the Institute for Health Metrics and Evaluation, "Age-adjusted firearm homicide rates in the US are 19 times greater than they are in France, and 77 times greater than in Germany. The US has 33 times the rate of firearm homicide seen in Australia" (Sirull, 2023). Violent crime is a clear issue throughout the country, and it would be important to highlight any suggested potential relationships with certain variables. For example, the Harvard School of Public Health looked into the relationship between gun ownership and homicide. Researcher David Hemenway and his team found that "where there are more guns there is more homicide" across both high-income nations and US states (Hepburn & Hemenway, 2004, 417). Along with gun ownership rates, here are 3 of the key questions that we will answer throughout this project:

Do states that legally allow capital punishment have a lower rate of violent crime per capita?

Do states with more police funding (in dollars per capita) have a lower rate of violent crime per capita?

Do states with higher gun ownership rates have a higher rate of violent crime per capita?

In addition to these questions, we will also be exploring the relationship between our response variable of violent crime per capita and numerous other potential explanatory variables including: median income, poverty rate, state political affiliation, the legality of abortion, and the decriminalization of marijuana.In a recent poll, nearly 2/3 of Americans said they believe crime is a serious problem (Smith, 2024), and by identifying variables we want to create meaningful inferences that can be put to action. It is important to note that as an observational study, we cannot draw any causation from the relationships we discover in this project. Still, strong relationships could be helpful to identify areas for future research.

Methods and Analysis

First in the model we added quantitative variables and using t-tests determined that all four of our qualitative variables were significant for the model with significance level of 0.15. We tested multicollinearity in the model using VIF and found the highest individual value to

be MI with 5.021 and the model average to be 2.93. Since no individual is over 10 and the average is less than 3 we decided multicollinearity is not a concern for our model. We then tested for qualitative variables with t-tests and determined none of the binary qualitative predictors to be significant. Using a nested F-test for the death penalty we determined that it was not significant with significance of 0.15. We proceeded to check for interactions in the model and using interaction plots for qualitative x qualitative interactions and found that there was an interaction between Abortion and the Death Penalty since there was significant crossing over of the levels. Testing the interaction with a nested F-test showed the interaction was not significant and therefore should not be included in the model. In labeled scatter plots, there was no evidence of qualitative x quantitative interactions. The final model only included the four quantitative variables. We then assessed model assumptions. The QQ-plot slightly deviated from linear and was slightly clustered in the middle while the histogram of residuals also deviated slightly from normality. Our regression model should be robust from these minor violations. The residual vs. fitted plot showed no fanning pattern or lack of fit so equal variance and lack of fit were not violated. Time series data was not used so independence was not violated. In checking for influential observations, the studentized-residuals plot identified 2,24,31, and 34 as influential. Cook's distance showed 2,24, and 31 as influential. Although there were overlapping influential observations, we did not remove them as we only have 50 observations and do not wish to shrink the data set. Our questions also address nationally and removing certain state data would not help provide a comprehensive answer. Our additional technique we implemented was the k-folds cross validation. This creates a machine learning model that resamples our data into k number of folds (subsets) to split the data into. Then, k-1 subsets are used to train the model and the other subsets are used as a validation set (GeeksforGeeks, 2021). For our model, we used a k value of 10 to split our data into 5 even groups, which produced an adjusted R-squared value of 0.455. This is marginally better than our original adjusted R-squared value of 0.379. By using the k-folds cross validation method, we obtained a more reliable estimate of the model's performance compared to a single train-test split (GeeksforGeeks, 2021). This helps in assessing how the model generalizes to new data and provides insights into its stability and robustness.

Results

Our final model is: $CrimeRate = -1034 + .6501PoliceFunding + 5.071GunOwnership + 44.56PovertyRate + .00514MedianIncome$. The adjusted R-squared value is 0.379. This means that about 38% of the variation in the violent crime rate is explained by our model. This is not particularly high, however for a complex phenomenon such as violent crime, it

is difficult to create a model which thoroughly explains the whole thing. Additionally, the global F-test on 4 and 45 degrees of freedom results in a test statistic of 8.465 which has a p-value of 0.000035, meaning that our model is adequate.

Conclusions

To answer our original research questions with our model, we look at the betas for the relevant explanatory variables.

For capital punishment we found that the death penalty had no significant impact on violent crime rates. The nested F-test had a p-value of 0.84 which is much higher than a significance level of 0.15.

For police funding, we find that increased police funding is correlated with increased violent crime rates. For a one dollar increase in police funding per person, we see a .6501 increase in the violent crime rate. This could potentially be explained by states that have high violent crime spending more to attempt to combat it.

For gun ownership, we find that increased gun ownership is correlated with increased violent crime. The rate of violent crime changes by 5.071 per percentage increase in gun ownership. This means that higher gun ownership is correlated with increased violent crime. This could be explained in multiple ways, such as people owning more guns to commit crimes with, or owning guns to protect themselves from crime.

A potential use for this model is predicting how violent crime rates might change when a new governor is elected for a state. A change to the governor's party combined with changes in policy such as legalizing marijuana, abortion, and the death penalty, and also proposed policy changes to increase police funding and reduce poverty, while holding median income and gun ownership constant, could be analyzed with this model. However, since the R squared of our model is not high, it may not actually be a useful model for prediction.

There are some improvements we can make to our model. Generally, violent crime is a very broad and nuanced topic. It might be more accurate and more applicable to narrow down our response variable to a specific type of violent crime, such as gun violence. As a result, we can find variables that have more of a direct effect on the response. Additionally, our model covers very contemporary issues, so it is important that we have the most recent data to reflect the current political, social, and economic conditions in each state.

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description
Violent Crime	Crime rate	Violent crime is composed of four offenses: murder and nonnegligent manslaughter, foricble rape, robbery, and aggravated assault. These are all crimes that involve force or threat of force. The rate is the number of reported offenses per 100,000 people in the population of the state. For example, a crime rate of 400 in a state with a population of 1,000,000 means that 4,000 violent crimes were reported.
Presence of the Death Penalty	DP	The death penalty is the highest punishment administratable by courts and means that the convicted will be put to death. Not all states allow this punishment, so there are two levels of this binary variable: 0 for no death penalty, 1 for death penalty.
Average police funding	PF	Police funding measures the amount of money a states spends on police departments per capita(person) in dollars.

Variable Name	Abbreviated Name	Description
Median household income	MI	The median household income is the 50th percentile of the household incomes for households in the United States, calculated using $2,500 intervals. Household income measures the amount of money that the providers of a household bring in.
Poverty rate	PR	The percentage of people living below the poverty line in each state. The poverty line depends on the number of individuals in a household, beginning at $14,580/year for individuals
Gun ownership rate	GO	The percentage of people who own guns in each state.
State Govenor's party	GP	A state's primary political affiliation, based on the party of the govenor. The levels of the qualitative variable are 0 for democrat and 1 for republican
The legality of abortion	LA	Whether abortion is legal in a state or not. Abortion is illegal in a state if it is completely banned. The levels of the qualitative variable are 0 for not legal and 1 for legal

Variable Name	Abbreviated Name	Description
The legality of marijuana	LM	Whether marijuana is legal in a state or not. Marijuana is considered legal in a state if it is decriminalized and allowed to be sold. The levels of the qualitative variable are 0 for not legal and 1 for legal

Appendix B: Data Rows

	State	Crime_Rate	Governor_Party	Death_Penalty_Present	Police_Funding
1	AL	409.1	R	Y	296
2	AK	758.9	R	N	555
3	AZ	431.5	R	M	401
4	AR	645.3	R	Y	244
5	CA	499.5	D	M	564
6	CO	492.5	D	N	402

	Gun_Ownership	Abortion_Legality	Poverty_Rate	Marijuana_Legality	Median_Income
1	55.5	No	16.3	No	59910
2	64.5	Yes	10.8	Yes	89740
3	46.3	Yes	12.9	Yes	73450
4	57.2	No	16.0	No	53980
5	28.3	Yes	12.3	Yes	85300
6	45.1	Yes	9.7	Yes	89930

Appendix C: Tables and Figures

Final model F-test and R squared

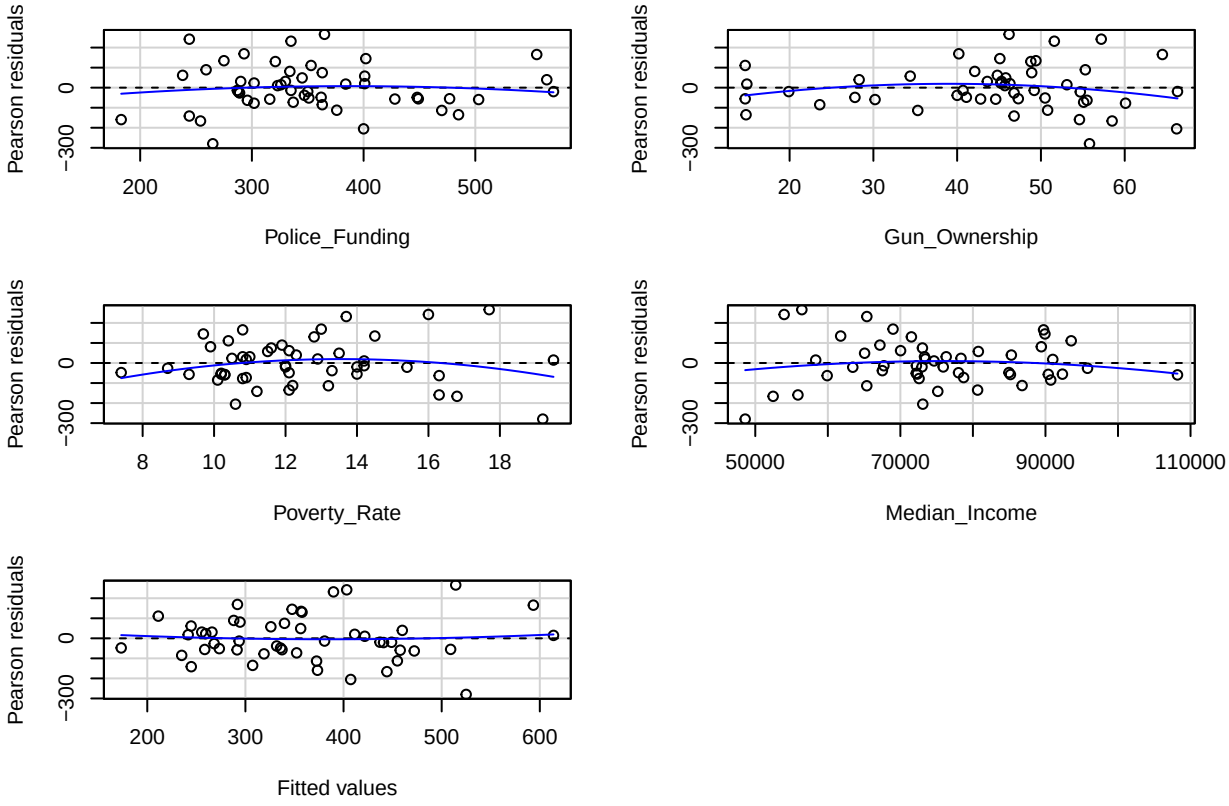
```
Call:
lm(formula = Crime_Rate ~ . - State - Marijuana_Legality - Governor_Party -
    Abortion_Legality - Death_Penalty_Present, data = Crime_Recompiled)

Residuals:
    Min      1Q  Median      3Q     Max
-279.94  -58.83  -16.48   60.53  266.17

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.034e+03  3.726e+02  -2.776 0.007993 **
Police_Funding  6.501e-01  2.454e-01   2.649 0.011085 *
Gun_Ownership   5.071e+00  1.594e+00   3.182 0.002649 **
Poverty_Rate    4.456e+01  1.228e+01   3.630 0.000721 ***
Median_Income   5.111e-03  2.989e-03   1.710 0.094144 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 118.3 on 45 degrees of freedom
Multiple R-squared:  0.4294,    Adjusted R-squared:  0.3786
F-statistic: 8.465 on 4 and 45 DF,  p-value: 3.515e-05
```

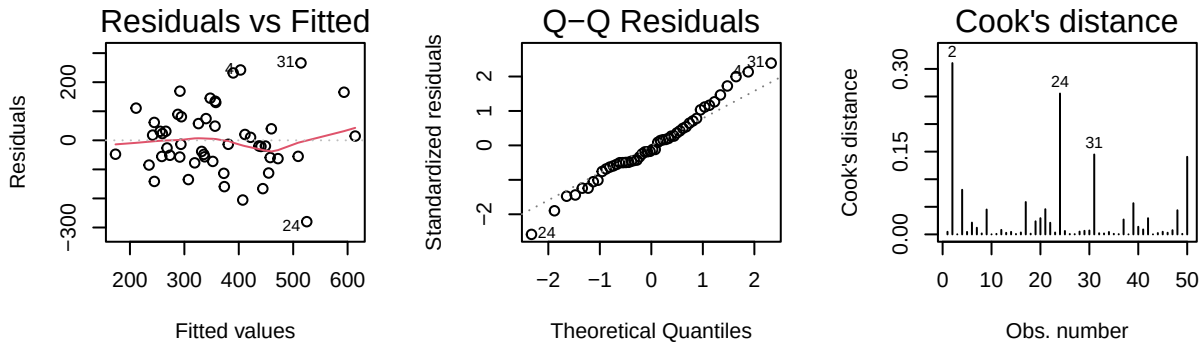
Residual Plots



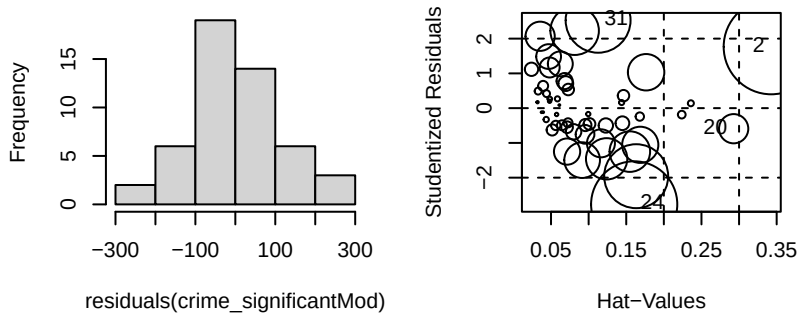
Assumptions and outliers plots

Influence plot values

	StudRes	Hat	CookD
2	1.7658007	0.3427015	0.31052164
20	-0.5915547	0.2928147	0.02940349
24	-2.7682889	0.1605641	0.25535419
31	2.5283378	0.1125962	0.14486034



ogram of residuals(crime_signific



K-fold analysis

```
Linear Regression

50 samples
 9 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 45, 46, 43, 45, 45, 43, ...
Resampling results:

  RMSE      Rsquared   MAE
  115.9898  0.4348683  92.734

Tuning parameter 'intercept' was held constant at a value of TRUE
```

Appendix D: References

Background

References Hepburn, L., & Hemenway, D. (2004). Firearm availability and homicide: A review of the literature.

Aggression and Violent Behavior: A Review Journal, 9(417), 40. https://www.hsph.harvard.edu/hicrc/firearms-research/guns-and-death/

Sirull, R. (2023, October 31). On gun violence, the United States is an outlier. Institute for Health Metrics and Evaluation. Retrieved March 28, 2024, from https://www.healthdata.org/news-events/insights-blog/acting-data/gun-violence-united-states-outlier

Smith, C. (2024, January 11). Why It's Confusing to Know Whether Crime's Really Up or Down. Why It's Confusing to Know Whether Crime's Really Up or Down. Retrieved March 28, 2024, from https://www.governing.com/urban/why-its-confusing-to-know-whether-crimes-really-up-or-down

Data

CBS News. (April 14, 2022). Gun Ownership by State 2022 [Graph]. In World Population Review. Retrieved March 28, 2024, from https://worldpopulationreview.com/state-rankings/gun-ownership-by-state

Center for Politics. (November 9, 2023). Party control of governorships following 2023 elections [Graph]. Retrieved March 28, 2024, from https://centerforpolitics.org/crystalball/2024-governor/

Center for Reproductive Rights. After Roe Fell: Abortion Laws by State [Graph]. Retrieved March 28, 2024, from https://reproductiverights.org/maps/abortion-laws-by-state/

Death Penalty Information Center. States with and without the death penalty – 2023 [Graph]. Retrieved March 28, 2024, from https://deathpenaltyinfo.org/states-landing

Defense Information System Agency. (March 1, 2024). MARIJUANA LEGALITY BY STATE - Updated March 1, 2024 [Graph]. Retrieved March 28, 2024 from https://disa.com/marijuana-legality-by-state

FBI. (October 16, 2023). Reported violent crime rate in the United States in 2022, by state (per 100,000 of the population) [Graph]. In Statista. Retrieved March 28, 2024, from https://www.statista.com/statistics/200445/reported-violent-crime-rate-in-the-us-states/

US Census Bureau. (September 13, 2023). Median household income in the United States in 2022, by state (in 2022 U.S. dollars) [Graph]. In Statista. Retrieved March 28, 2024, from https://www.statista.com/statistics/233170/median-household-income-in-the-united-states-by-state/

US Census Bureau. (October 18, 2022). Police and Corrections Spending by State [Graph]. In MoneyGeek. Retrieved March 28, 2024, from https://www.moneygeek.com/living/state-policing-corrections-spending/#detailed-findings

US Department of Agriculture. (June 16, 2023). Percent of total population in poverty, 2021 [Graph]. Retrieved March 28, 2024, from https://data.ers.usda.gov/reports.aspx?ID=17826

Supplemental Code and Analysis Help

GeeksforGeeks. (2021, December 28). K-fold Cross Validation in R Programming. Retrieved April 21, 2024, from https://www.geeksforgeeks.org/k-fold-cross-validation-in-r-programming/#