The Roots of Crimes

Chris Jiang

4/7/20

Table of Contents

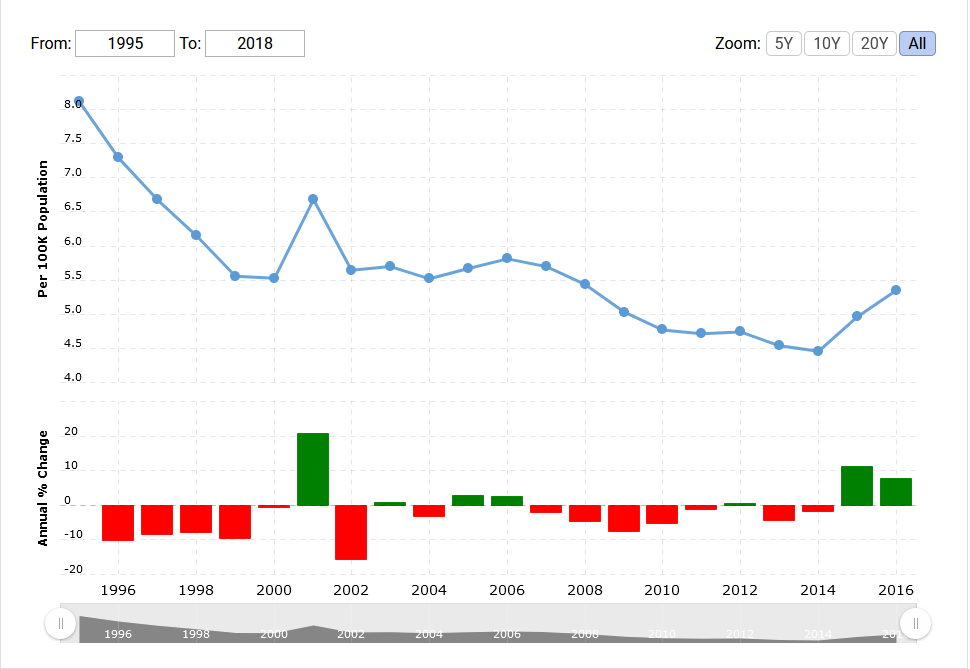
* **Page 3:** Introduction
* **Page 3:** Objective
* **Page 4:** Data Set
* **Page 4:** Data Manipulation
* **Page 14:** Conclusion

**Introduction**

Crime will always be an issue that plagues the world and unfortunately it will never go away. Everyday we hear reports of robberies and murders and in way, has become a norm to us a society. Public perception matters and continuing in this way will definitely have negative consequence Being able to solve the puzzle of crime will help improve the public perception and help alleviate their woes. To be able to solve this problem, we need to first better understand why crimes occur by looking for a relationship between crime and multiple variables.

**Objective**

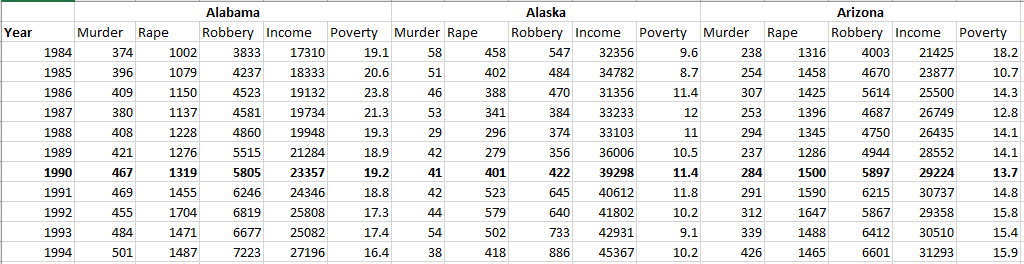
The main goal of this research is to determine if there are any variables that move with the crime rate such as population, income etc. to help better understand the movement of the crime rate. Solving or getting a better understanding of the different moving parts can help us focus on other goals that lower crime rate as a byproduct while focusing on other impactful social issues. The crime rate has fallen overall, but in order to better understand the reason for the fall in crime, we can compare the movement of the crime rate with other different key issues.



**https://www.macrotrends.net/countries/USA/united-states/crime-rate-statistics**

**Data Set**

The data set that will be used was founded on a crime reporting website that shows and lists the amount of crimes per year.



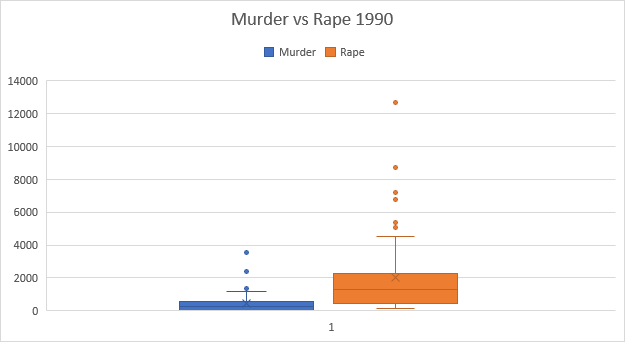
The variables that will be focused on are:

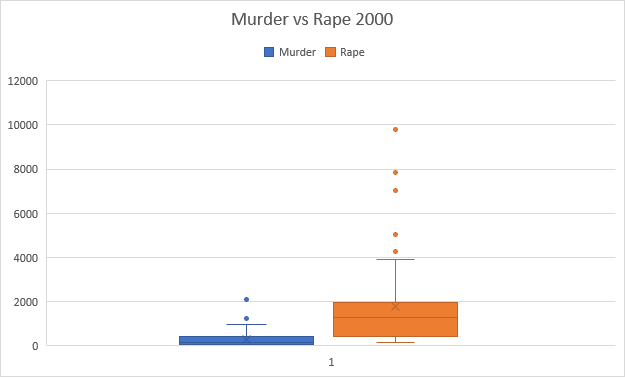
* Year
* States
* Murder
* Rape
* Median Income
* Poverty Rate

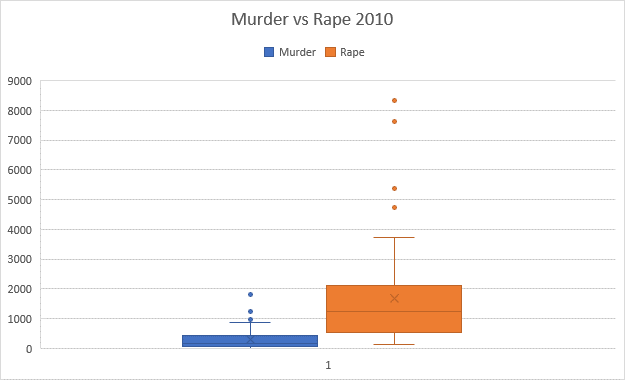
The original barebones data set only listed the year, states, and the crimes. The rest of the information was then compiled and then added to the original data set. Each of the variables will be shown in a time series where the movements of them will be tracked and determined if there is a strong or weak relationship between each of them.

**Data Manipulation**

To begin with, we will be focusing on three specific years, 1990, 2000, and 2010. For these three years, we will compare each of the two crimes as a box plot. Box plots are good as a visual in determining the range of the data set that is being worked with and also shows in a visual sense if there any outliers attached to your data. Below are the three box plots for each year:





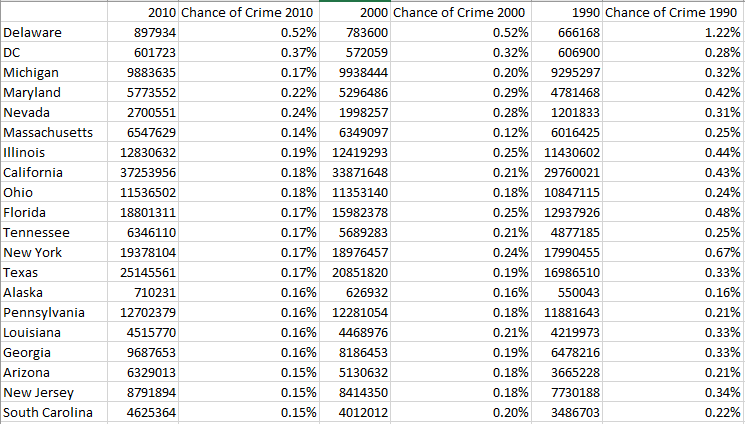


As you see above, we have 3 separate boxplots that show the variation of the data set is being worked with. When comparing them, the range decreases over time. The 1990 box plot has the max rape value between 12,000 and 14,000 and the 2010 box plot has a max value between 9,000 and 9,000. In regards to murder, 1990 has a max value between 2,000 and 4,000 and 2010 shows that the max value of murder is between 1,000 and 2,000. This shows that as years go by, the crime rate has fallen for each of the different states.

Another way to get a better look at the different spread of the data is by using quadrants. The 50 states are divided into 4 regions and be compared with 2 of the more serious crimes, murder and rape. The quadrants will be divided with the medians for rape and murder for that specific year. Through this we will be able to visualize where each of the different states stand:

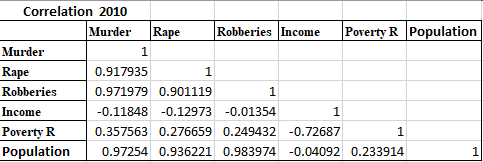
The quadrants to look out the most for are the ones in the top right, due to that quadrant meaning that the state has higher rapes and murders when compared to the medians. The majority of the southern region, compared to the three other regions, are in the upper right of the quadrants. As shown in the charts as well, California and Texas are in the upper right areas as expected, due to the population of those States being extraordinarily high. We will be focusing on Texas, California and the 2 lowest population states, Wyoming and DC based on three years’ worth of population data. Murders, rapes and robberies were added together as a new variable, crime and we will compare that value to the values of poverty and median income to examine if there any relationships for each of the 4 states. Results might differ for each state.

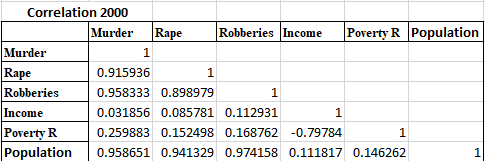
As you can see above, one would expect that the lower the population, the lower the amount of crimes. Texas, California and Wyoming support this general idea but DC is questionable. District of Columbia has the second lowest population in the United States but in terms of crimes, it hovers between 4000 and 5000 in 2014. That might seem like a lot, but in the previous quadrant chart, you can see that in terms of murder vs rape, it’s on the bottom left area which means that it is lower than the median for both murder and rape for that current year. To better understand this, we will be taking account population to determine the crime rate based on population for each state. Some of the results are shown below but I will point out the numbers that stand out:

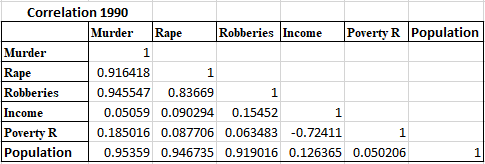


This table has a new variable, **chance of crime.** This variable takes account population into determining violent crimes, by dividing the total number of violent crimes by the population, and then sorted based on highest chance of crime. The values under the year columns are the population for that current year. In the year of 2010, the states with the highest chance of crimes are Delaware and DC. Both states have a fairly low population and yet the chance of a crime is so high compared to other states. The two lowest states in 2010 are Vermont and Wyoming with population of 625741, 563626 in 2010 respectively.

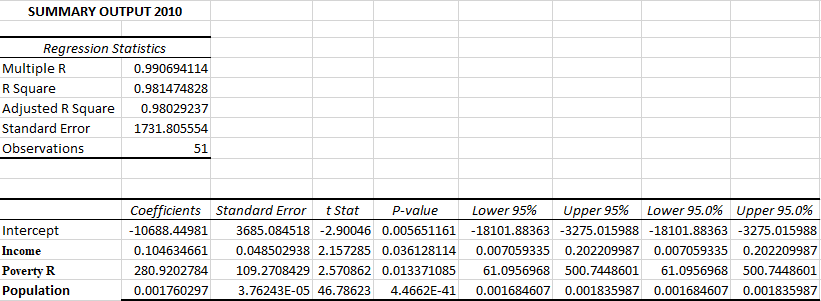
To further investigate this to determine if this is true, we will determine the correlation coefficients for each of the crimes, median income, and poverty rate for each of the year and then they will be compared with each other. We will be focusing on 3 years again, 2010, 2000, and 1990 due to having the population information. Results are show down below:

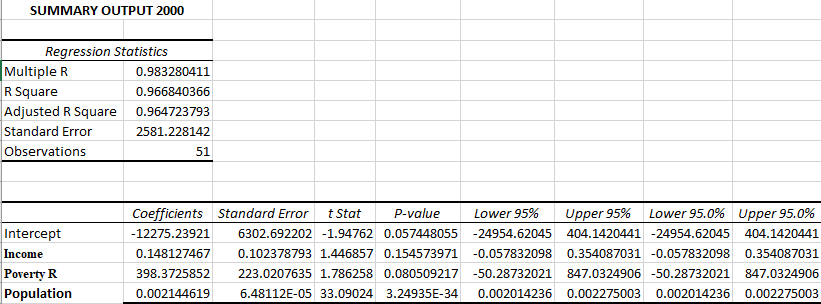


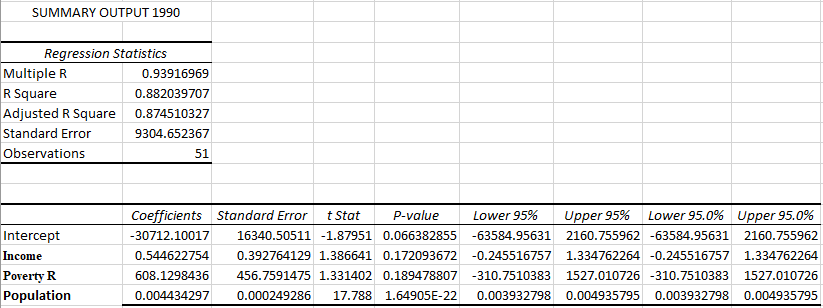




Interestingly the results for the correlation analysis show that population has a very high correlation coefficient in regards to the different crimes. For each of the separate years, the coefficients are greater than 90%. Income and Poverty Rates have very low correlation with murder. Median income is the lowest and the highest coefficient for Poverty is about .35 and the lowest is .18. The coefficient increased as the years go by, so it might be that poverty is becoming more of an issue compared to the past. The correlation coefficient for Income also becomes negative in the year of 2010, which means that as income goes up, the crimes decrease. It’s interesting how this wasn’t the case for the other 2 years. Now that we have the correlation coefficients, we are going to determine if the variables are significant. The different crimes were combined into a new variable as **violent crime.** Results are show below for each year:

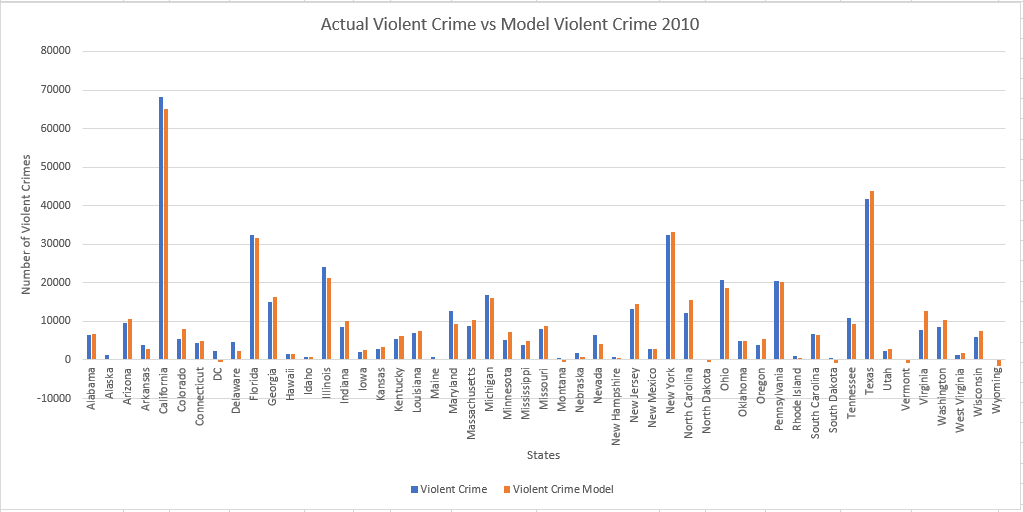


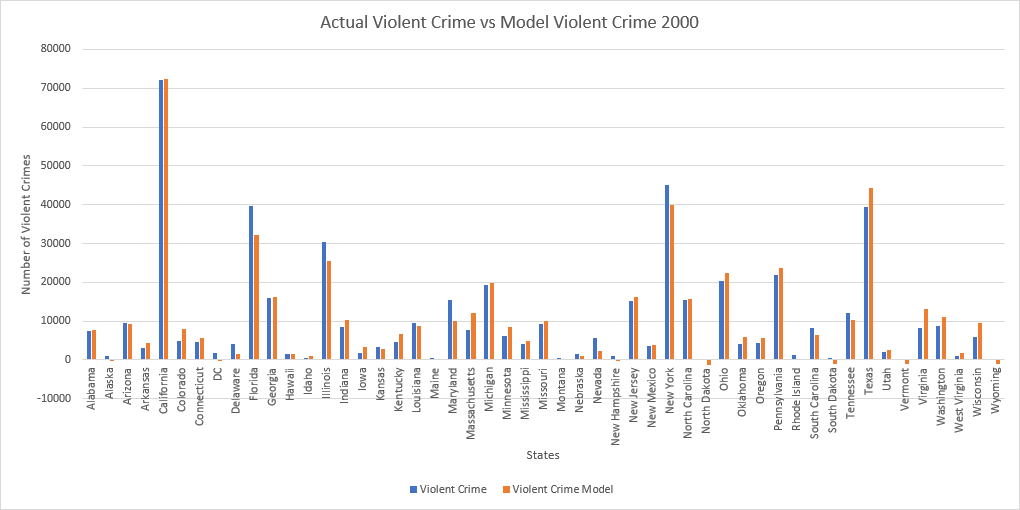


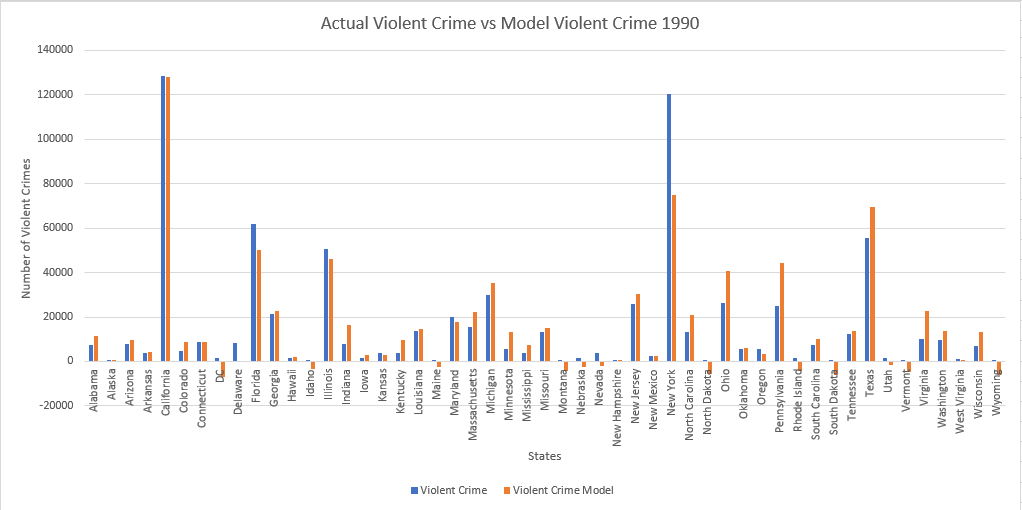


Looking at the data, we see that the R2 for each of the models is very high, between .88 and .98. The value of population also seems to be very consistent as well, being a significant value for all 3 years. This isn’t the same for Income and the Poverty Rate however. For year of 2010, those two variables were significant in the 95% level but not in 2000 and 1990. This data does support with what we have the results of the correlation data.

Below is a visualization comparing our predictions against the actual amount of violent of crimes for each of the years:

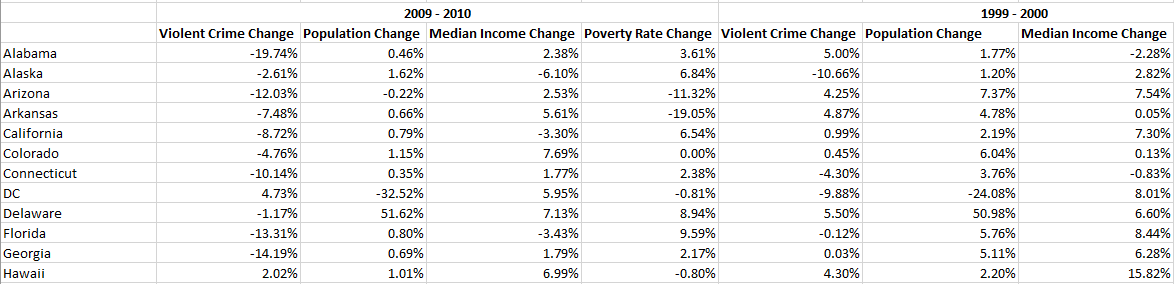




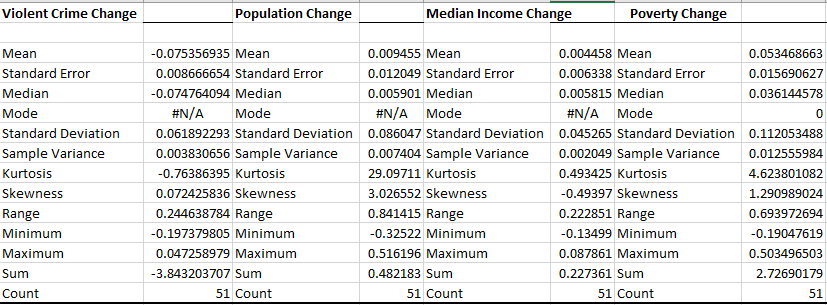


As you can see, the results are pretty accurate which makes sense with such a high R2 value. Some predictions show a negative number which isn’t accurate at all. Those predictions are usually for the states with very low population but for the states with high population, it is generally pretty accurate.

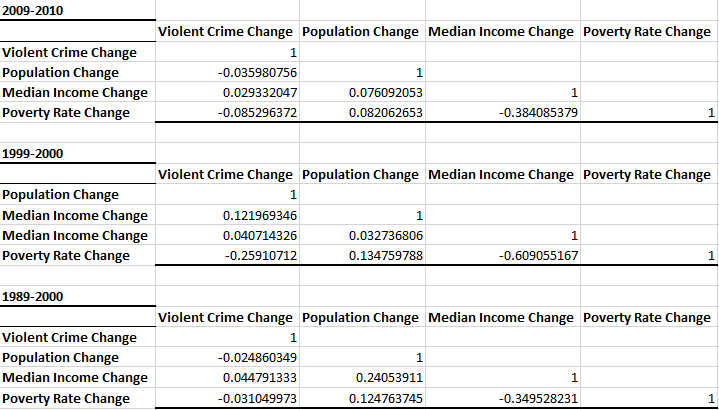
Last but not least, our final study will be based on the significance of the yearly crime changes when compared to the yearly changes of income, poverty rate, and also population. To do this, a new table was created and part of it is shown below:



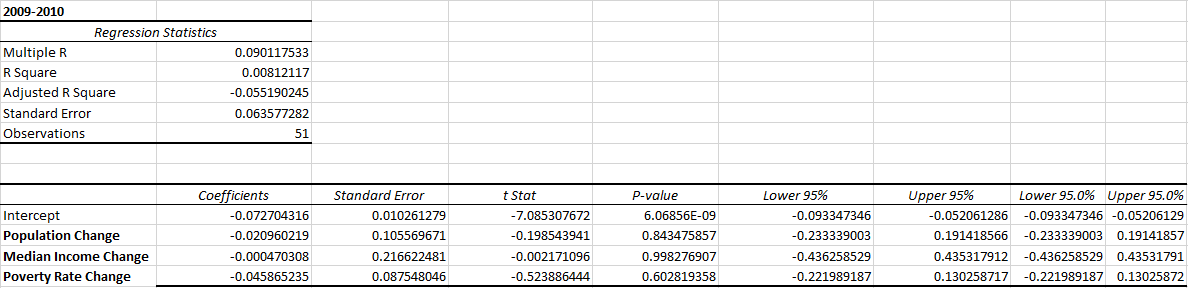
I obtained the changes in percent for 3 years, 2009-2010,1999-2000, and 1989-1990 and calculated the percent change for each of the years. A descriptive statistics test was also run on the data above to determine normality and the values for skewness and kurtosis came back relatively small:



Correlation analysis is then run on the data to demine if any relationships exists between each other:



The correlation for the data seems to not be as strong as the previous data set. Looking at the population change versus the violent crime change, the relationship between the two seems fairly weak. That is the same for all the other variables when compared to the change in violent crime. Last but not least, a regression analysis will be used to see if the variables are statistically significant.



Above is the regression analysis for 2009-2010 but all results are roughly similar. None of the percent changes are significant in predicting the percent change in crime rates unfortunately. The R2 is very small and the p-values for each of the variables are greater than .05 so unfortunately, it is not significant at the 95% level.

**Conclusion**

In the research, it is shown that the variable of population has a very strong relationship with the amount of violent crimes in the different states. However, just because the correlation coefficient is high doesn’t mean that the states will have a high chance in crimes. When taking account population and the crimes, we notice that some states have high chance of crimes even with low populations such as DC and Delaware. Predicting crime changes is roughly difficult when only looking at changes in population, median income, and poverty as shown the relationship is roughly weak and the p-values very high. Last but not least, this is my results from the study and this information isn’t concrete and 100%. Not all crimes are documented and reported and there are plenty of crimes out there that were never noticed. More study needs to be done on this topic to further understand the relationships between crimes and the cause of it.