Socioeconomic Factors that Predict Crime Rate

Jessica Macchione, Sanjay Tiwari, Kendrick Turner, Jana Shepard,

Melissa Kearney, Gabrielle Naimie and Tyler Ouellet

Harvard University Extension School

**Introduction**

While extensive research has been conducted on the causation of crime, the spectrum of theories in this area is broad. Numerous theories suggest that poverty predicts crime; but, crime is a complex phenomenon and its causes are, in many cases, multi-dimensional (Barkan & Rocque, 2018; Fergusson, Swain-Campbell & Horwood, 2004). Understanding the predictors of crime and the underlying elements of criminal behavior are essential to identifying safe neighborhoods and communities. The present study was designed to evaluate whether certain socioeconomic factors can effectively predict crime (or the lack thereof). This research evaluates socioeconomic data collected by the United States Census, the Federal Bureau of Investigation’s Uniform Crime Reporting Data and the United States Law Management and Administrative Statistics to determine which demographic or socioeconomic factors are significant predictors of crime. The goal of the research is to provide a regression model of all significant variables that predict violent crime.

The study of criminology indicates that criminal predictability is comprised of a social component and an individual component. Research indicates that the sociological component has a greater bearing on criminal predictability than the individual element (Murray, Irving, Farrington, Colman, Bloxsom, 2010). The most common sociological factors that provide a ripe environment “to create” a criminal are family and community. Importantly, these factors are multi-layered. For instance, the familial influence on the potential future criminality of an individual depends on a number of elements. Simply growing up in a single parent family isn’t predictive of future criminal conduct. However, change that dynamic to a female-led single parent family, at poverty level, with an element of substance abuse and limited parent interaction and the risk factor escalates exponentially (Musick, 2002). Conversely, growing up in an impoverished two-parent family that is engaged with the child is negatively correlated with future criminal behavior (Murray, Irving, et. al, 2010). Familial elements that have proven to be predictive of criminal behavior include mother-led single parent families, children born to teenage parents, and limited or non-existent parental engagement. Such familial elements should be evaluated both individually and systematically to ascertain their impact – parent engagement cannot be undervalued (U.S. Census Bureau, Table C2).

The other primary element of the sociological theory is neighborhood and community. While community poverty on its own initially suggests fertile ground for criminal activity, there are underlying elements that are more predictive of the opportunity for criminal behavior. Some impoverished communities are challenged with lower quality education and healthcare, some of those same communities have strong cultural bonds and established codes of conduct. In neighborhoods where communities have well-established bonds and expectations, the cultural thread of the community trumps the damning economic influence (Sampson, Raudenbush & Earls, 1997). In addition to community elements, the racial makeup of the community and the presence of law enforcement also have a correlation with crime (Barkan & Rocque, 2018). It is suggested that the neighborhoods that have higher rates of crime have fewer law enforcement agents (Levitt, 2004). Further, neighborhoods with a minority of Caucasians may have less access to law enforcement (Barkan & Rocque, 2018).

In harmony with the research discussed above, the researchers predicted that variables that describe family life, education, poverty, racial dispersion and community would be highly correlated with predicting the rate of violent criminal activity. The specific variables from the dataset of interest include: single family status, divorce rate, law enforcement makeup and racial makeup.

**Method**

The present study was conducted using existing data from the United States Census, the United States Law Management and Administrative Statistics (LEMAS), and the Federal Bureau of Investigation’s Uniform Crime Reporting (USR) data. These surveys include socioeconomic data from distinct communities (n=1933) across the United States.[[1]](#footnote-1)

Researchers used Statistical Package for the Social Sciences (SPSS) to conduct all statistical outputs to produce two different models. The data set was obtained by means of Internet search. Researchers determined that regression analysis would be used to predict per capita crime. Researchers took a hierarchal and forced entry approach in determining which variables would best predict violent crime in a multiple regression model. For each entry, the researchers ran statistical tests to test the assumptions of regression analysis. Those tests included a plot of the data, K-S test and PP Plot to establish normality. Tolerance and variation influence factor (VIF) statistics were ran to test multicollinearity. The Durbin-Watson test was ran to test for autocorrelation. Researchers conducted Levene’s test was to test for homogeneity of variance. The researchers also plotted the data against the regression models. Each of these tests was employed to assure that the data fit the assumptions of regression analysis, which include: linear relationship, normality, heterogeneity of residuals, no autocorrelation and homoscedascity.   
  
We chose to look at these three variables to investigate level of education. This breaks up education into three levels. Those with less than a 9th grade education, those without a high school diploma, and those with a bachelor’s degree or higher education.

PctLess9thGrade

PctNotHSGrad

PctBSorMore

Next, we looked at families with 2 or more parents. This was broken down with four variables. All households with 2 parents, households in housing projects with 2 parents, households with young kids that have two parents, and households with teens that have 2 parents.

PctFam2Par

PctKids2Par

PctYoungKids2Par

PctTeen2Par

Next, we looked at the total divorce rate within the community.

TotalPctDiv

Next, we looked at the racial breakdown of each community based on percent black, white, Asian and Hispanic.

racepctblack

racePctWhite

racePctAsian

racePctHisp

Finally, we looked at the number of fulltime police officers in the community and the number that actively patrol a neighborhood.

LemasSwornFT

LemasSwFTFieldOps

We then used these variables to predict the violent crimes in a population:

ViolentCrimesPerPop

then ran simple correlations for the aforementioned variables. The researchers took note of the variables whose absolute value were greater than .5. The variables that remained were racepctblack, racepctwhite, totalpctdiv, pctfam2par, pctkids2par, pctyoungkids2par, pctteen2par. These variables were forced in and out of the model to provide the most statistically presented model, based off of their statistical significance and fit.

**Results**

Initially researchers determined that he percentage of white population (b=-.31, p<.01), percentage of two parent families (b=-.96, p<.01) and number of police officers sworn in field operation (b=-.21, p<.01) were significantly negatively correlated with violent crimes. Overall, the model accounted for a significant amount of the sample (r=.79, p<.01) and the population alike (r2=.62, p<.01). Overall, the model that the analysis yielded was predicted crimes=-.30 (percentageofpopulationwhite) -.7 (percentageoftwoparentfamilies) -.22 (numberofpoliceofficerssworninfieldoperation) +1.15 (see appendix 4). Upon further inquiry, the team realized that number of sworn field opereation (LEMASWFTFIELDOPS) could be biasing the model because the information was only available to 319 communities, thus limiting the data set to those 319 communities out of a set of 1,993. The team repeated the process, throwing out the law enforcement data to provide a better fit.

The final output determined that the percentage of white population (b=-.48, p<.01), percentage of two parent families (b=-.76, p<.01) and percentage of population divorced (b=.20, p<.01) were significantly negatively correlated with violent crimes. Overall, the model accounted for a significant amount of the sample (r=.78, p<.01) and the population alike (r2=.61, p<.01). The regression equation that the analysis yielded was predicted crimes=-.48(percentageofpopulationwhite) -.76 (percentageoftwoparentfamilies) +.20 (percentagedivorced) +.78. The final model Pearson R-square deviance from sample to population was marginal, indicating our final model is free from overfitting and given sample size of N=1994 and model explaining 61% variance, we can be confident on these findings.

These variables for the most part met the assumptions of linearity (see plots), normality (histogram reveals a close to normal distribution SD=.98). All the variables tested showed tolerance scores greater than 0.1 (tolerancepercentagewhite=.44, tolerancepercentagekidswith2parents=.44, tolerancenumberofpolicofficerssworninfieldoperation=.91). The results suggest there could be an issue of interrelated predictor variables as pctpopblck (b=-.79, p<.01) and pctpopwhite (b=-.79, p<.01) roughly seemed to explain the same variance when the simple correlation was made to the other (see appendix 1). That is that the percentage of black population equally influenced the percentage of white population and vice-versa. In addition, the overall model proved to be a good fit with a Durbin-Watson score between 1.5 and 2.5 (DW=2.02). There is an issue of heteroskedastity when looking at the residual plots of our final model. The variance is not even for all levels of our output against the residuals of the model (see appendix 1). The data was transformed to control for this factor and other normality issues. The transformed model, yielded roughly the same results as the final model, indicating that these issues could not be controlled for.

**Discussion**

Consistent with past research, the present research supports the notion that home life and racial dispersion of a community are important predictors of a safe neighborhood. In addition, the higher percentage of divorce was positively correlated with crime rate within a neighborhood. Each of those factors accounted for a significant amount of the model in predicting crime. In the overall scope of things, the results suggest that a neighborhood with high rates of two parent families and higher proportions of Caucasians is likely to be correlated with less crime, while a neighborhood with a high divorce rate would likely be correlated to higher crime rates.. Given our results, it is important to consider how the demographic makeup of communities impact safety. A family looking for a safe neighborhood would be able to use this model, with the end goal of finding a safe neighborhood.   
 While there is reason to conclude that the present model is effective, the study did not go without limitations. First and foremost, according to the website from which the data was obtained, the dataset was submitted on July 13, 2009, which was based off of data obtained from 1995 (UCI 2009). Thus, the most obvious shortcoming is that the present data reflect trends relative to the mid 1990s in the United States, and may not be applicable to the year of 2018. A more up to date data set could yield different predictive variables that are more relevant to current trends.   
 In addition, as outlined in the results section, while the overall model was a good fit for regression analysis, there were some problems with normality. The fact that the percentage of black residents and percentage of white residents explained roughly the same variance is problematic, seeming to suggest that those two predictor variables are correlated. In addition, the unfortunate fact that the law enforcement variables had to be thrown out does not necessarily signify that law enforcement could be a significant predictor of crime in a community as it proved to be in the initial model. In order to grasp a better understanding of the relationship between the rates of law enforcement in the community and crime, it would have been ideal to have law enforcement data for all 1,993 communities. As presented, the current data set did not allow us to include that variable in the final analysis, which has been supported by previous research.

The present research present correlations to predictive variables and predicted crime within amongst communities. The adage, “correlation does not equal causation” is applicable to this research. Some reading this research may be tempted to conclude that these factors “cause” less crime, but would be mistaken as the model only states that the relationship between these factors and crime are significant.

In that same breath, the current model only accounts for one layer of a certain type of predictors for crime. Only socioeconomic factors were used as predictive variables. It’s possible other layers (such as biological) were not accounted for. That is to say that the present research does not answer the question as to why or if there is a significant interaction between socio-economic factors and crime. The current results could suggest that there could be specific tactics that Caucasians and or two parent families do differently, compared to divorced families, and families of other ethnicities. Conversely, the data could suggest that there are different law enforcement tactics depending on the race being policed. The data could also indicate a portion of both of these variables could account for crime rates within a community. However, the present data set for this analysis does not include the necessary information to yield a multivariate analysis (MANOVA) in order to compare the possible differences of child rearing amongst Caucasians and other racial counterparts, or the possible differences of child rearing in two parent homes compared to divorce homes or the possible different policing tactics differences between races. Another up to date dataset that contains that information to those would be necessary to understand crime at other layers in relation to the aforementioned socio-economic factors. It follows logical sense that if there are high rates of different policing methods and or parenting practices, that those high rates could explain the level of crime within a community. Researchers recognize that the answers to these questions could be uncomfortable, but possibly necessary to help understand the casual elements of criminal behavior on both the individual and communal level.

For the aforementioned reasons, the researchers encourage readers to view the present research as a component of criminal behavior, but discourage readers from using the present research as conclusion to a conclusive explanation to criminal behavior. The researchers regard the present research as complimentary and supportive to other research. Overall, more research and analysis is needed to help the general population understand how crime works on a communal level.

References

Barkan, S.E. & Rocque, M. (2018). Socioeconomic status and racism as fundamental causes of street criminality. *Critical Criminology, 1-21*. Retrieved from http://doi.org/10.1007/s10612-18-9387-x

Fergusson, D., Swain-Campbell, N., & Horwood, J. (2004). How does childhood economic disadvantage lead to crime? *Journal of Child Psychology and Psychiatry, 45:5*, 956-966. doi: 10.111/j.1469-7610.2004.t01-1-00288.

Gatti, U. & Verde, A. (2012). Cesare Lombroso: Methodological Ambiguities and brilliant intuitions*. International Journal of Law and Psychiatry. 35 (1): 19-26*. doi: 10.1016/j.ijlp.2011.11.004

Jung, H., Herrenkohl, T., Klika, J., Lee, J., & Brown, E. (2015). Does child maltreatment predict adult crime? Reexamining the question in a prospective study of gender differences, education and marital status. *Journal of Interpersonal Violence, 30(13)*, 2238-2257. doi:10.1177/0886260514552446

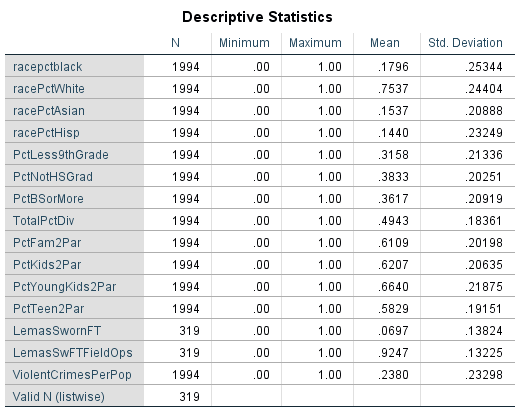
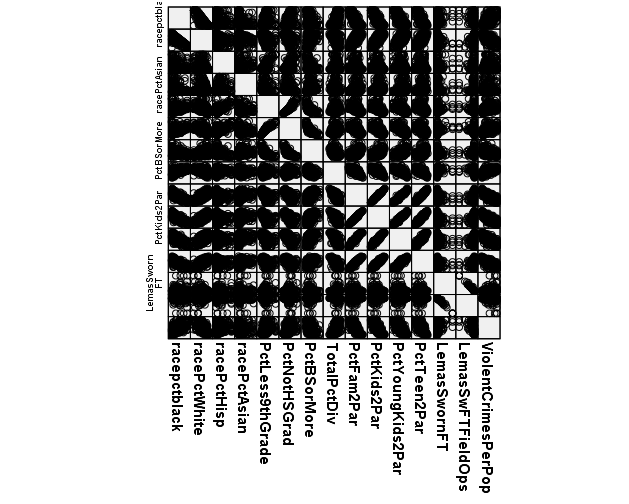
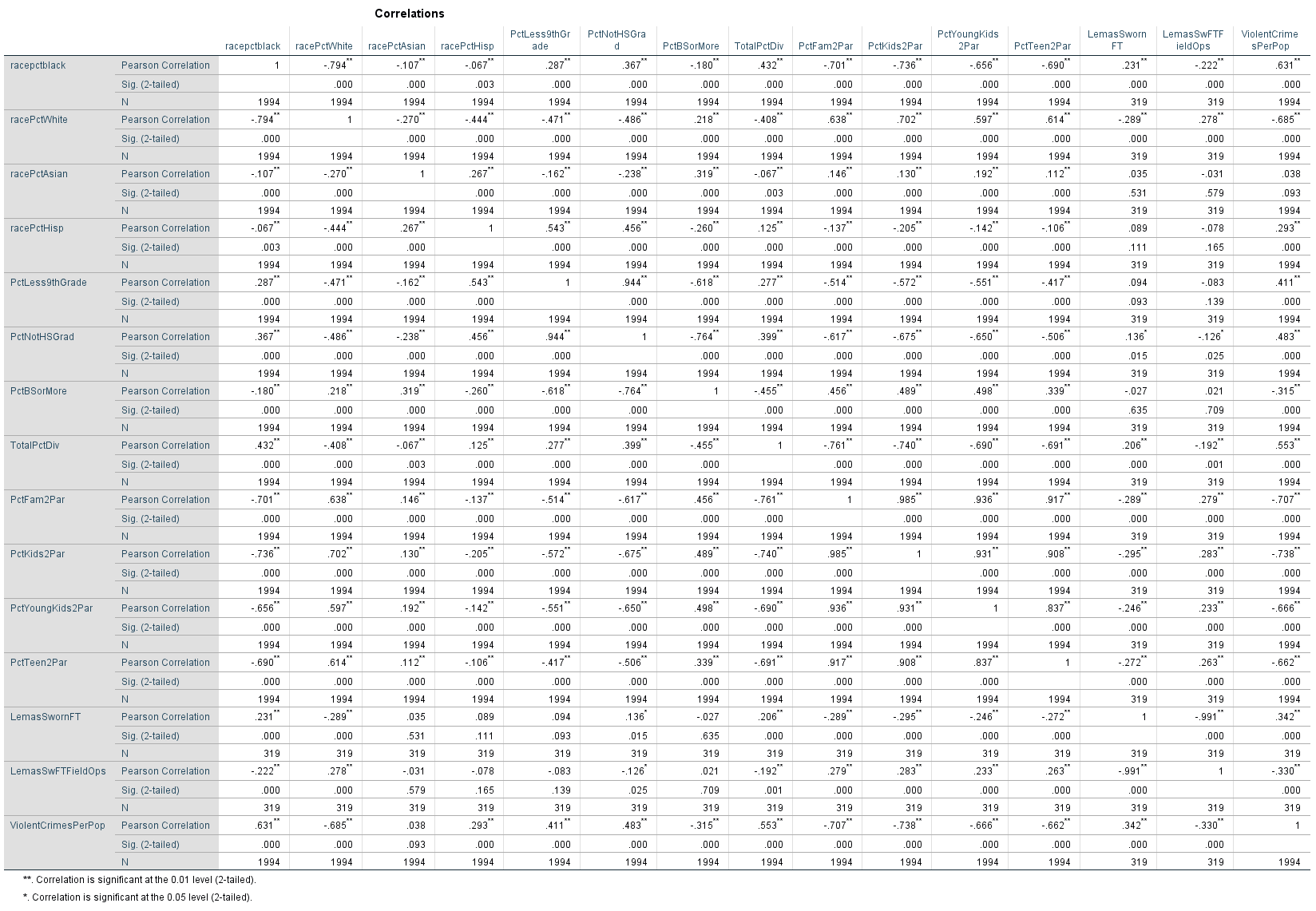
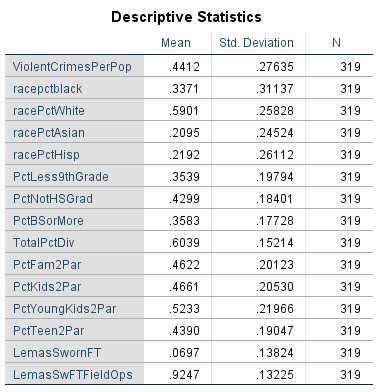
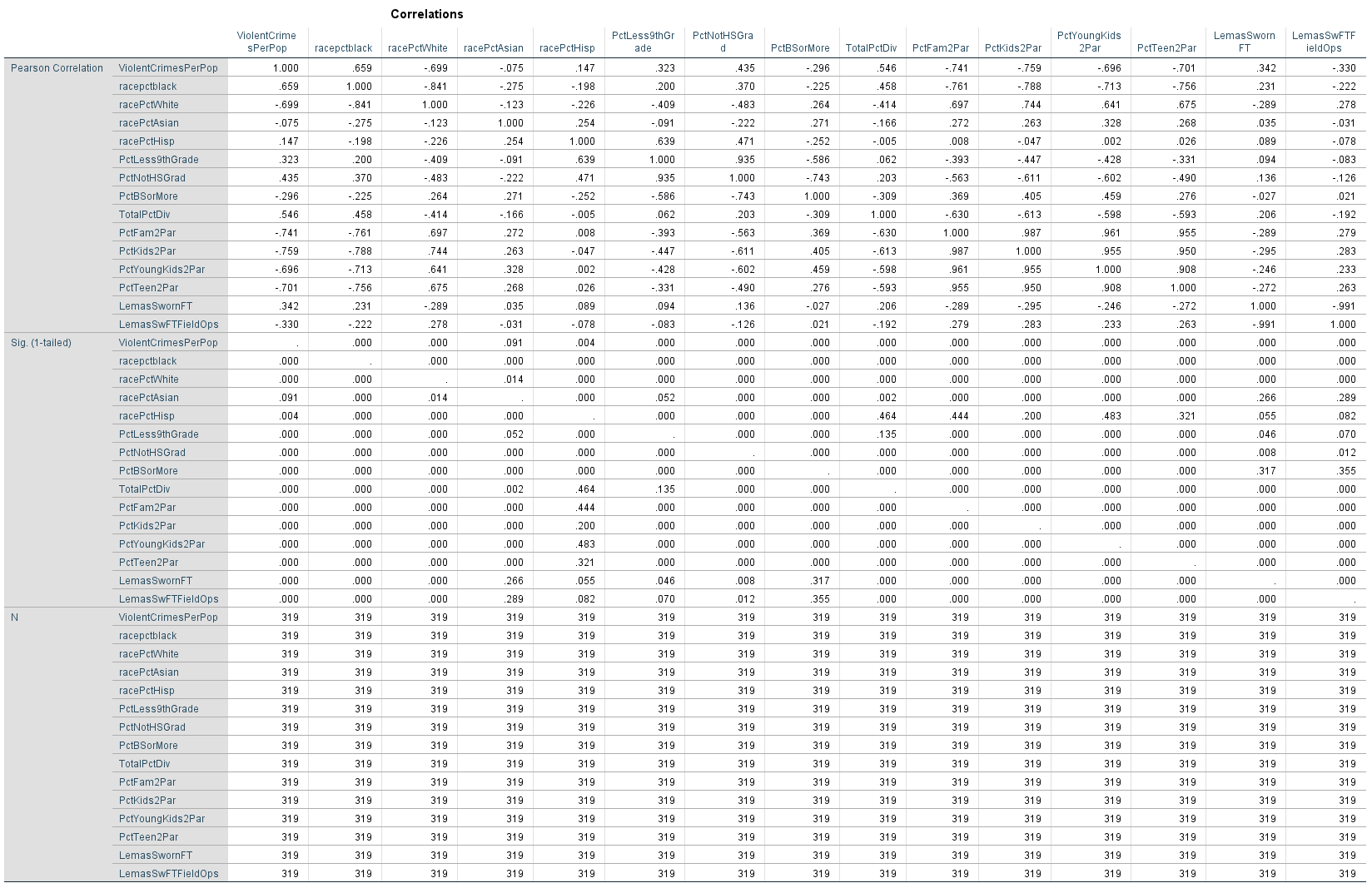
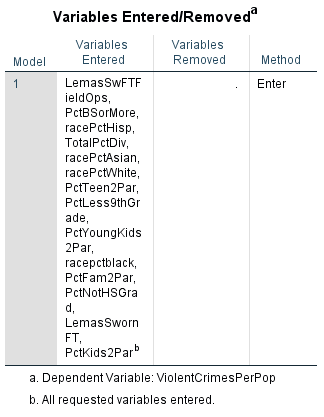
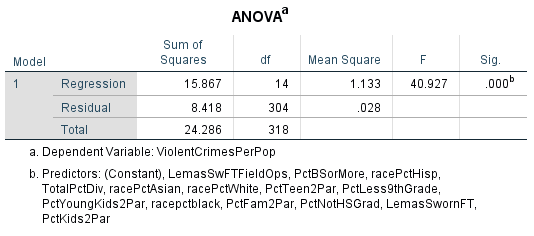
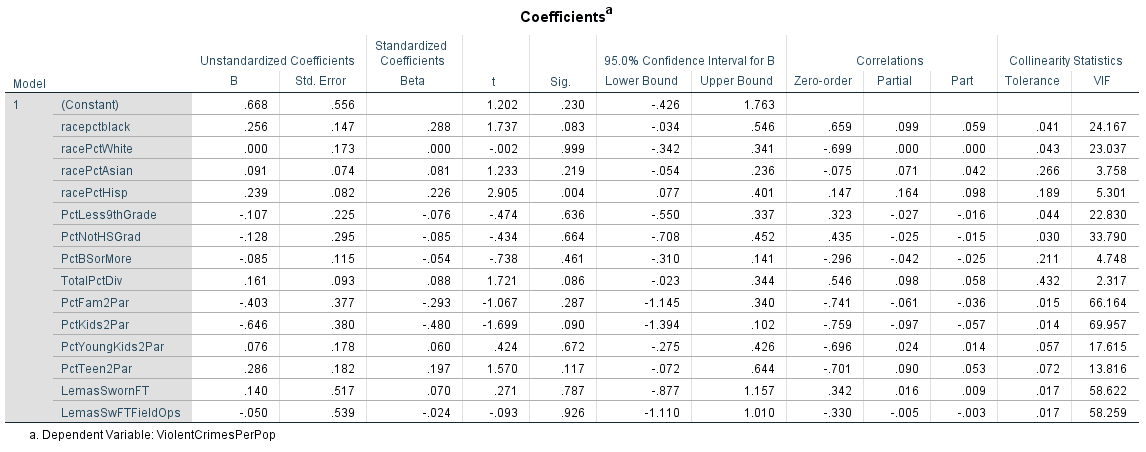
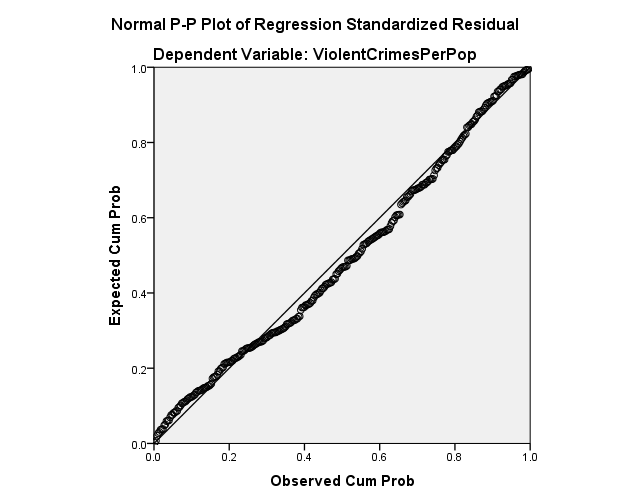
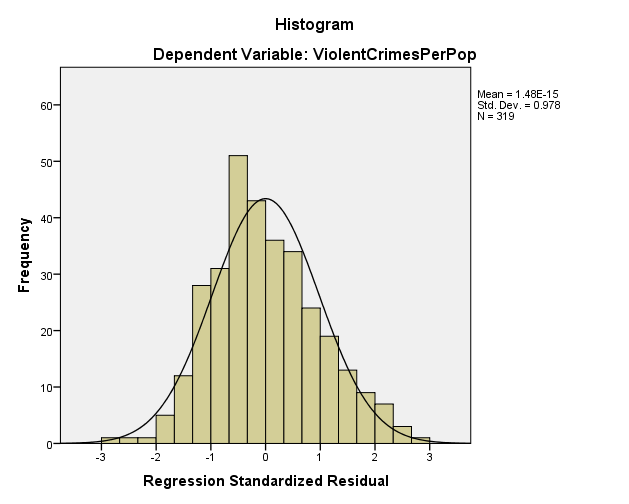
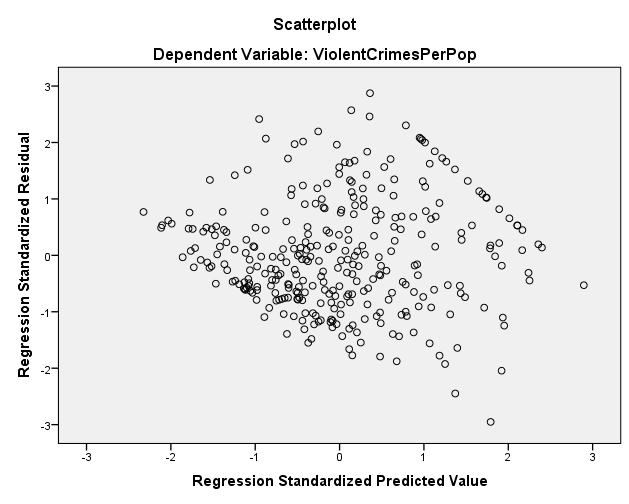
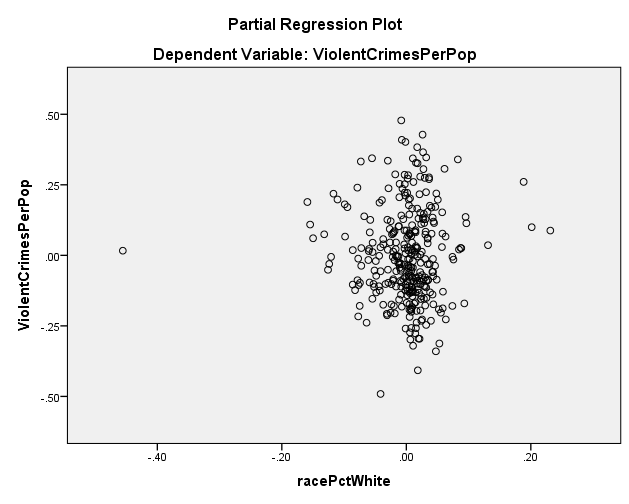
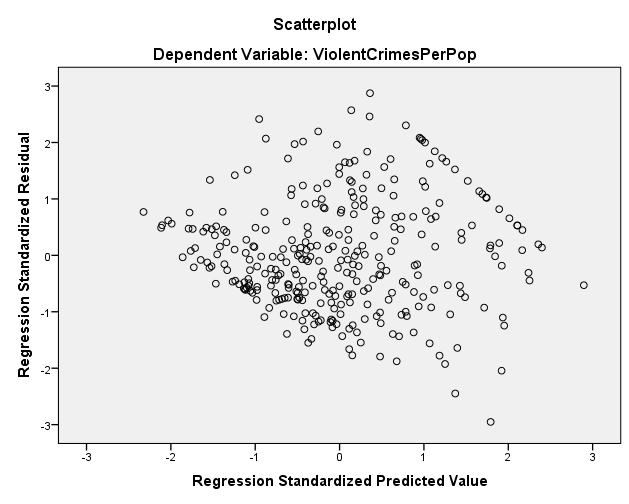
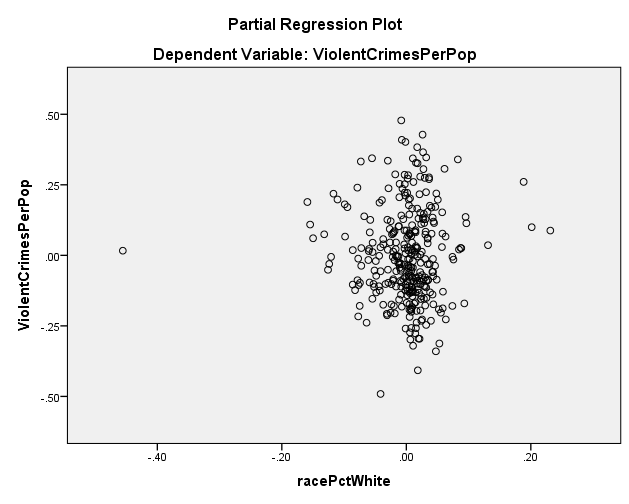
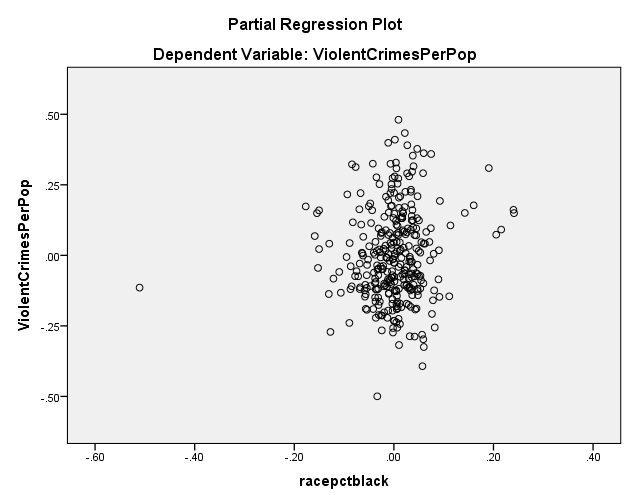
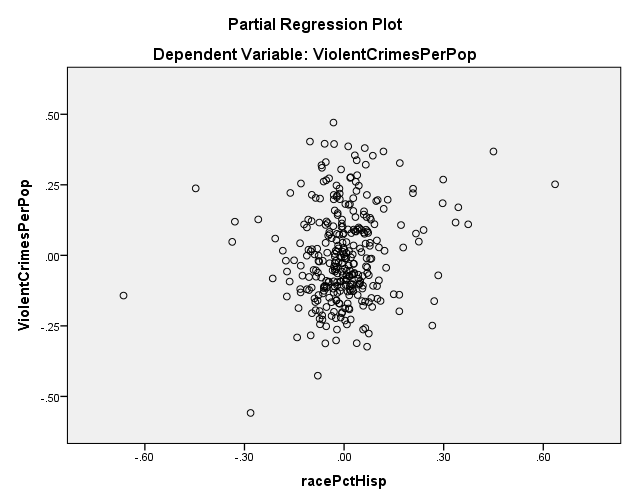
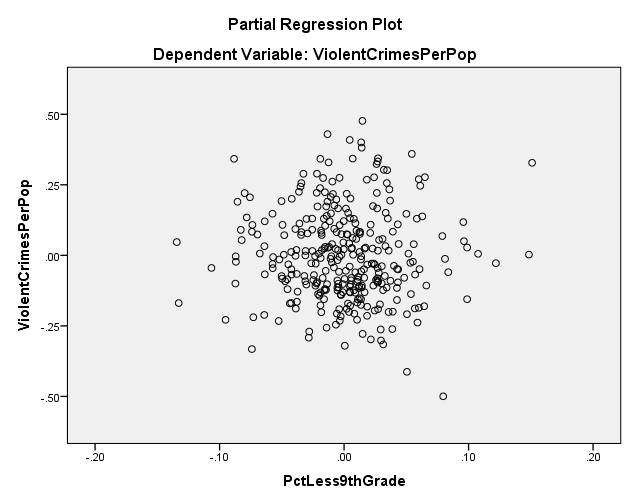
Levitt, S. (2004). Understanding whey crime fell in the 1990s: four factors that explain the decline and six that do not. *Journal of Economic Perspectives, 18:1*, 163-190. doi: 10.1257/089533004773563485

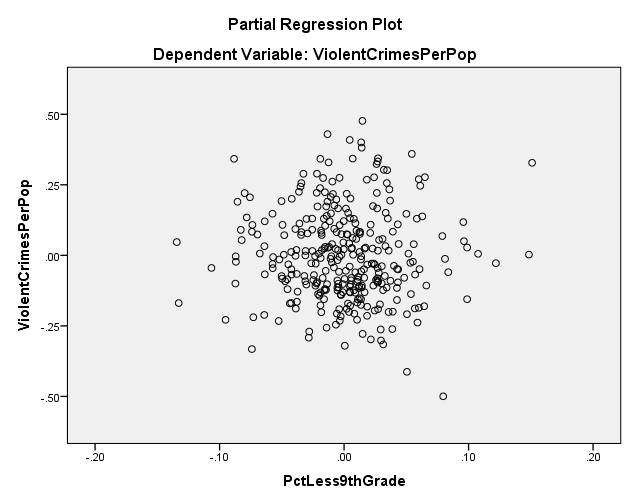
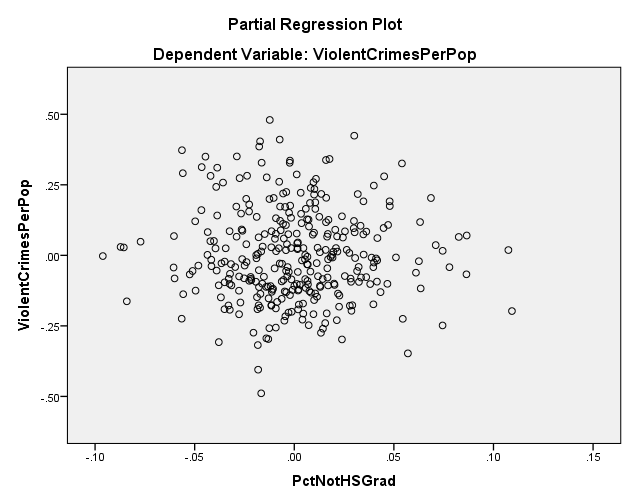
Minkov, M. & Beaver, K. (2016). A test of life history strategy theory as a predictor of criminal violence across 51 nations. *Personality and Individual Differences 97*, 186-192. doi: 10.1016/j.paid.2016.03.063

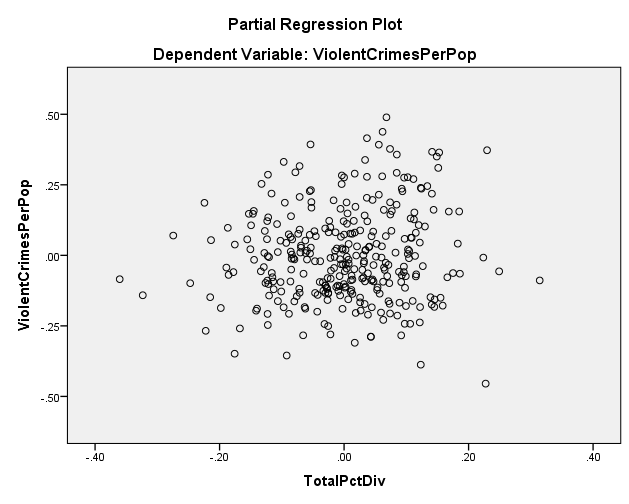
Murray, J., Irving, B., Farrington, D., Colman, I., & Bloxsom, A.J. (2010). Very early predictors of conduct problems and crime: results from a national cohort study. *Journal of Child Psychology and Psychiatry*, 51:11, 1198-1207. doi: 10.111/j.1469-7610.2010.02287.x

Sampson, R.J., Raudenbush, S.W., & Earls, F. (1997). Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science, 277*. doi: 10.1126/science.277.5328.

U.S. Census Bureau, Table C2. Household relationship and living arrangements of children under 18 years, by age and sex: 2013.

Appendix 1  
contains the summary statistics of the first iteration.  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  




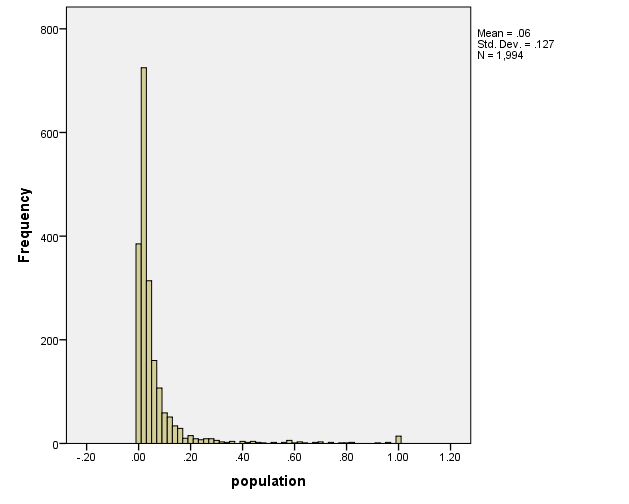
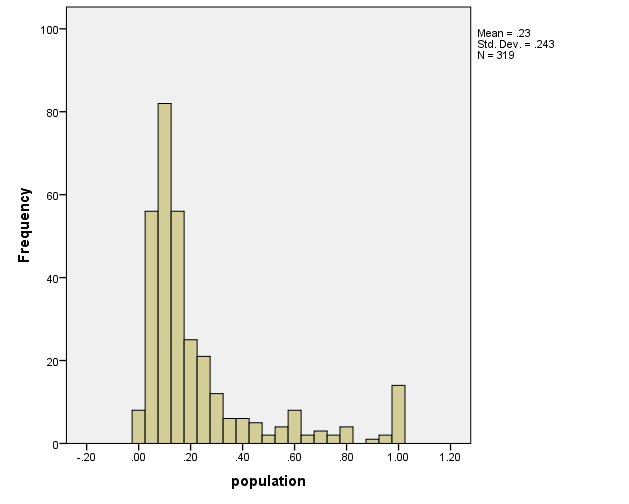
Appendix 2  
contains the statistics of the final model and the final model transformed

Table 1 (Final Model)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | |
| B | Std. Error | Beta | Lower Bound | Upper Bound |
| 1 | (Constant) | .784 | .029 |  | 27.160 | .000 | .727 | .840 |
| racePctWhite\_LN | -.476 | .030 | -.322 | -15.616 | .000 | -.536 | -.416 |
| PctKids2Par\_LN | -.762 | .046 | -.442 | -16.409 | .000 | -.853 | -.671 |
| TotalPctDiv\_LN | .197 | .038 | .106 | 5.141 | .000 | .122 | .272 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Correlations | | | Collinearity Statistics | |
| Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) |  |  |  |  |  |
| racePctWhite\_LN | -.677 | -.330 | -.220 | .467 | 2.140 |
| PctKids2Par\_LN | -.746 | -.345 | -.231 | .273 | 3.665 |
| TotalPctDiv\_LN | .544 | .114 | .072 | .469 | 2.134 |

(TRANSFORMED MODEL HERE)  
Appendix 3  
This appendix contains histograms showing the number of cases that reported law enforcement statistics.

TOTAL N=1997  
  
TOTAL REPORTED LAW ENFORCMENT N=319  


Appendix 4

This appendix contains the final model with the inclusion of law enforcement

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Correlations** | | | | | |
|  | | ViolentCrimesPerPop | racePctWhite | PctKids2Par | LemasSwFTFieldOps |
| Pearson Correlation | ViolentCrimesPerPop | 1.000 | -.699 | -.759 | -.330 |
| racePctWhite | -.699 | 1.000 | .744 | .278 |
| PctKids2Par | -.759 | .744 | 1.000 | .283 |
| LemasSwFTFieldOps | -.330 | .278 | .283 | 1.000 |
| Sig. (1-tailed) | ViolentCrimesPerPop | . | .000 | .000 | .000 |
| racePctWhite | .000 | . | .000 | .000 |
| PctKids2Par | .000 | .000 | . | .000 |
| LemasSwFTFieldOps | .000 | .000 | .000 | . |
| N | ViolentCrimesPerPop | 319 | 319 | 319 | 319 |
| racePctWhite | 319 | 319 | 319 | 319 |
| PctKids2Par | 319 | 319 | 319 | 319 |
| LemasSwFTFieldOps | 319 | 319 | 319 | 319 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryb** | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .791a | .626 | .623 | .16975 | .626 | 175.948 | 3 | 315 | .000 | 2.024 |
| a. Predictors: (Constant), LemasSwFTFieldOps, racePctWhite, PctKids2Par | | | | | | | | | | |
| b. Dependent Variable: ViolentCrimesPerPop | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 15.209 | 3 | 5.070 | 175.948 | .000b |
| Residual | 9.076 | 315 | .029 |  |  |
| Total | 24.286 | 318 |  |  |  |
| a. Dependent Variable: ViolentCrimesPerPop | | | | | | |
| b. Predictors: (Constant), LemasSwFTFieldOps, racePctWhite, PctKids2Par | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | | Correlations | | | Collinearity Statistics | |
| B | Std. Error | Beta |  |  | Lower Bound | Upper Bound | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 1.149 | .067 |  | 17.060 | .000 | 1.016 | 1.281 |  |  |  |  |  |
| racePctWhite | -.305 | .056 | -.285 | -5.487 | .000 | -.414 | -.195 | -.699 | -.295 | -.189 | .441 | 2.268 |
| PctKids2Par | -.696 | .070 | -.517 | -9.958 | .000 | -.834 | -.559 | -.759 | -.489 | -.343 | .440 | 2.275 |
| LemasSwFTFieldOps | -.220 | .075 | -.105 | -2.909 | .004 | -.368 | -.071 | -.330 | -.162 | -.100 | .910 | 1.099 |
| a. Dependent Variable: ViolentCrimesPerPop | | | | | | | | | | | | | |

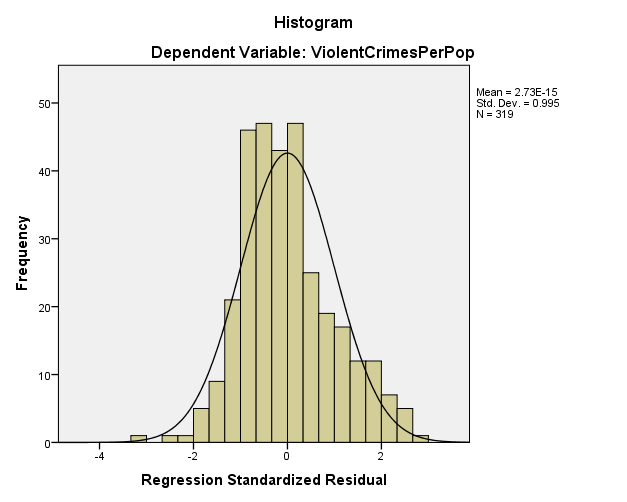
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Coefficient Correlationsa** | | | | | |
| Model | | | LemasSwFTFieldOps | racePctWhite | PctKids2Par |
| 1 | Correlations | LemasSwFTFieldOps | 1.000 | -.106 | -.118 |
| racePctWhite | -.106 | 1.000 | -.723 |
| PctKids2Par | -.118 | -.723 | 1.000 |
| Covariances | LemasSwFTFieldOps | .006 | .000 | -.001 |
| racePctWhite | .000 | .003 | -.003 |
| PctKids2Par | -.001 | -.003 | .005 |
| a. Dependent Variable: ViolentCrimesPerPop | | | | | |

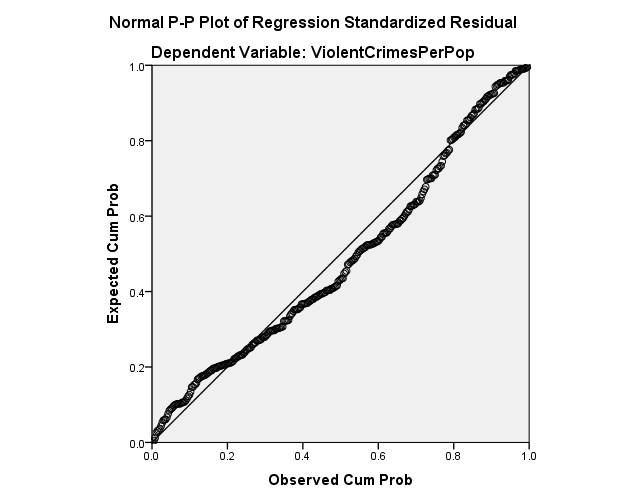
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Collinearity Diagnosticsa** | | | | | | | |
| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
| (Constant) | racePctWhite | PctKids2Par | LemasSwFTFieldOps |
| 1 | 1 | 3.813 | 1.000 | .00 | .00 | .00 | .00 |
| 2 | .136 | 5.294 | .04 | .13 | .13 | .03 |
| 3 | .041 | 9.620 | .00 | .86 | .86 | .00 |
| 4 | .010 | 19.671 | .96 | .00 | .00 | .97 |
| a. Dependent Variable: ViolentCrimesPerPop | | | | | | | |

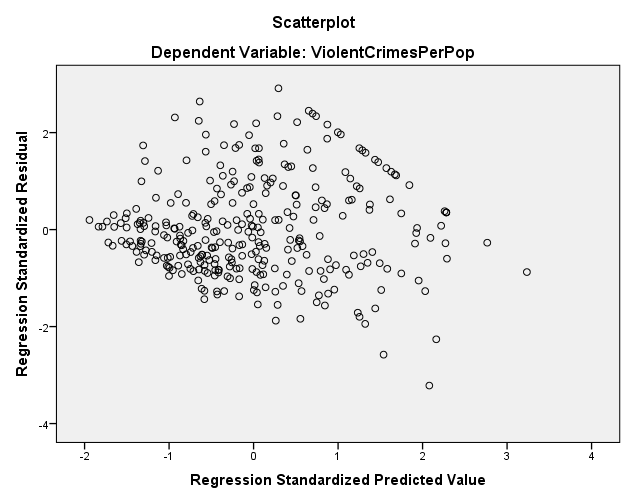
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Casewise Diagnosticsa** | | | | |
| Case Number | Std. Residual | ViolentCrimesPerPop | Predicted Value | Residual |
| 694 | -3.216 | .35 | .8960 | -.54597 |
| a. Dependent Variable: ViolentCrimesPerPop | | | | |

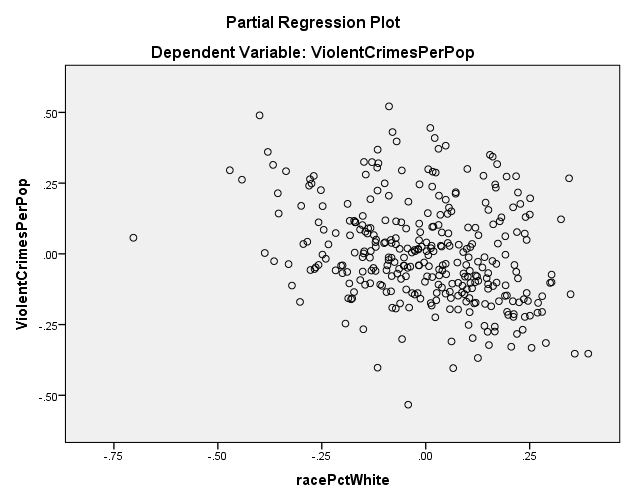
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Predicted Value | .0164 | 1.1485 | .4412 | .21870 | 319 |
| Std. Predicted Value | -1.942 | 3.234 | .000 | 1.000 | 319 |
| Standard Error of Predicted Value | .010 | .070 | .017 | .008 | 319 |
| Adjusted Predicted Value | .0158 | 1.1762 | .4414 | .21929 | 319 |
| Residual | -.54597 | .49444 | .00000 | .16894 | 319 |
| Std. Residual | -3.216 | 2.913 | .000 | .995 | 319 |
| Stud. Residual | -3.254 | 2.921 | -.001 | 1.003 | 319 |
| Deleted Residual | -.55873 | .49708 | -.00019 | .17157 | 319 |
| Stud. Deleted Residual | -3.305 | 2.956 | .000 | 1.006 | 319 |
| Mahal. Distance | .113 | 53.121 | 2.991 | 5.834 | 319 |
| Cook's Distance | .000 | .288 | .004 | .018 | 319 |
| Centered Leverage Value | .000 | .167 | .009 | .018 | 319 |
| a. Dependent Variable: ViolentCrimesPerPop | | | | | |

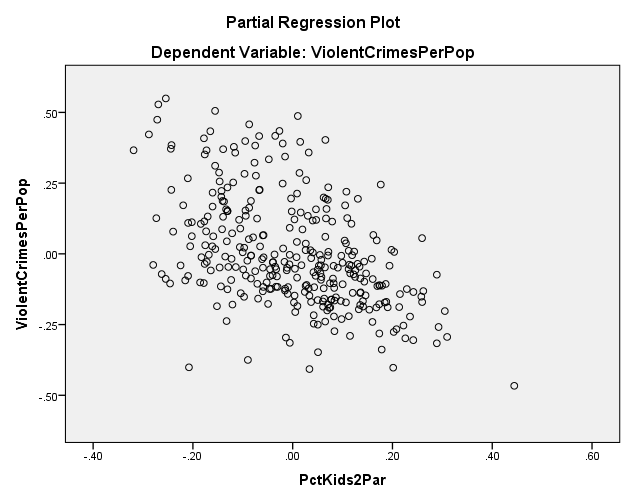
**Charts**

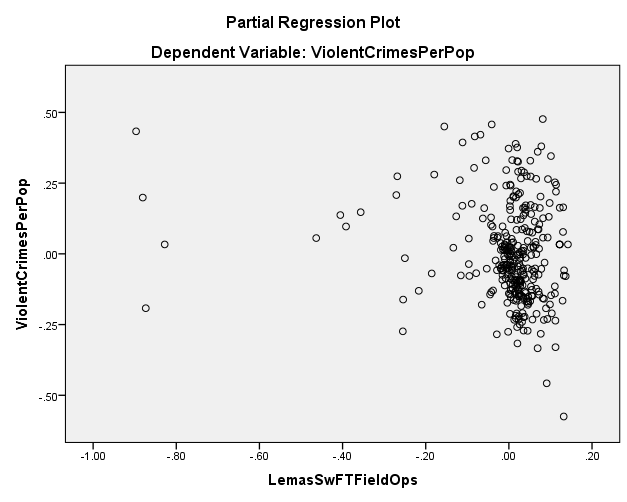












1. [↑](#footnote-ref-1)