**Statistics 2010 Project: Crime in the U.S.**

Kelly Yuson, Rudy Chiu, Seth Wagner, and Kendall Rasmussen

May 1, 2017

**1.1 Introduction**

Have crime rates in the United States changed over time? Constant news reports describing heinous crimes may give off the impression that the violent crime rate is hopelessly on the rise; however, a thorough analysis of FBI data on violent crime in the U.S. may reveal what may be counterintuitive to many: that the violent crime rate has been steadily decreasing for two decades, falling by nearly half.

**1.2 Methods**

To determine if there has been a statistically relevant trend in U.S. crime over recent years, it is of interest to compare the relationship between the year and annual crime rate. These variables of interest were manually inputted into SAS from the FBI’s data on their website. This input can be seen in appendix Figure A1. After entering the data, we used correlation and linear regression analysis to discover a relationship, as well as said relationship’s strength, between annual crime rate and time. All relevant SAS input can be found in the appendix and the resulting output will be discussed in detail in the following section.

**1.3 Results/Discussion**

The correlation procedure in SAS was run and the results are presented below in Table 1.



Table 1: Correlation data for Crime Rate in the US vs Year

Our data indicates that there is a -0.94491 correlation between our variables. This tells us that there is a strong, negative relationship between the crime rate and year, or in other words, there is strong evidence that the crime rate in the US has decreased over time. Visual evidence of this phenomenon can be seen in the regression plot below in Figure 1.



Figure : Scatterplot of Crime rate vs Year in the US with the regression line inserted.

From the above plot, we can see that the data fits a negative, linear pattern with an r-squared value of 0.8928. This means that the linear regression line given below explains approximately 89% of the variability in the model. The regression line is given by the following equation,

$$y=31567.9-15.47x$$

This equation tells us that as each year passes, the crime rate per 100,000 in the U.S. is expected to decrease by approximately 15.47. This equation can be used to predict the crime rate as a function of the year of interest.

 Because we are estimating the true slope of the population from the data, it’s important to acquire a 95% confidence interval for the true regression slope. From the table below, we can see that this confidence interval is (-12.81, -18.11). This means that we are 95% confident that the true unknown population slope, β, lies in this interval.



While our correlation appears to be very strong and while a considerable percent of the variability seems to be explained by our regression line, we must be wary. Shown above is our residual plot. Typically, we would like to see our residuals like white noise about the horizontal line. However, our residuals seem to follow a pattern. This suggests that a linear relationship between annual crime rate and year might not be the best; however, we will proceed with caution.



Table 2: Confidence Interval information for Linear Model Hypothesis Testing

Next, we further tested the hypothesis that there is not a linear relationship between annual crime rate and year. Our results are displayed above in table 2. Our null hypothesis (H0) was that there was no linear relationship and we are testing this against the alternative hypothesis (Ha) that there is a linear relationship,

$$H\_{0}: β=0$$

$H\_{a}:β\ne 0$.

We tested this at the α = 0.05 significance level. From the information in table 2, the t-statistic was determined to be -12.25, with a p-value less than 0.0001. This indicates that we would have less than 1 in 10000 chance of obtaining sample data producing a t-statistic this extreme, or more extreme, from 0 if the null hypothesis were true. Therefore, we reject our null hypothesis at the given significance level and conclude that there is some linear relationship in the data. In addition, because 0 is not contained in the 95% confidence interval for beta that we found earlier, we again have strong evidence that a relationship between annual crime rate and year exists.

**1.4 Conclusion**

The data presented clearly indicates a strong, negative linear relationship between annual crime rate and year in the United States. From this relationship, we can say that the crime rate has not increased over time, but rather it has decreased (good news for us)!!

**References**

Criminal Justice Information Services Division. (2013). Crime in the United States

by Volume and Rate per 100,000 Inhabitants, 1994–2013. Available from: <https://ucr.fbi.gov/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/tables/1tabledatadecoverviewpdf/table_1_crime_in_the_united_states_by_volume_and_rate_per_100000_inhabitants_1994-2013.xls>

**A. Appendix**

**A.1 SAS Code**

Yearly crime rate data from 1994 to 2013 was manually input into SAS,



Figure A1: SAS input for US crime data.

Correlation data was acquired by running the following input,

**proc** **corr** data = crime;

var rate year;

**run**;

Figure A2: Correlation procedure input for Crime rate vs Year

Linear Regression data was acquired by the following SAS input,



Figure A3: Regression procedure input for Crime rate vs Year

Group Contributions:

Kelly- Wrote the project proposal, SAS Code, reviewed and edited the final paper

Rudy- Helped write the SAS Code, wrote the project interim report, reviewed and edited the final paper

Seth- Helped find a dataset, reviewed project proposal, code, and interim report, wrote the rough draft of the final paper, formatted the final paper

Kendall- Wrote the intro and conclusion of the final paper, reviewed the final paper