Draft: Violence Symposium, 3rd November 2011

Modeling violent crime in The Bahamas

Yan Lyansky

School of Mathematics, Physics and Technology
College of the Bahamas

Abstract

Our goal is to find the best predictors of the violent crime rate (violent crimes divided by population). We assume that the government will be able to change policy to lower the crime rate if it knows the determining factors that influence crime. We compare unemployment, GDP, and population to the violent crime rate using linear models. We proceed to investigate non-linear and multi-variable models. Only population and GDP were found to be related to crime; increases in population increased crime, while increases in GDP decreased crime.

Background

Crime has been an escalating problem in the Caribbean. *In Crime in the Caribbean: Provisional Evidence* by McElroy and Roccanti (2005) question the sustainability of economic systems based on tourism given the large increase in crime. In The Bahamas, the general public perceives crime as a problem that it is out of control. Each week newspapers run murder, armed robbery, and/or stabbing stories. The number of murders continues to grow as well as the level of gun crime (Hanna, 2011). Criminals are becoming more brazen with their armed robberies as evidenced by the First Caribbean Bank assault (Smith, 2010) and an eco-tour attack (Thompson, 2009). According to sentiments expressed on local radio and television, it is clear that the Bahamian people believe it is time for a solution before things get worse. The police commissioner is under pressure to find a solution to the problem (Black, 2011). The purpose of this study is to examine macro-level factors which may influence crime in The Bahamas, and so allow government to set policies which would be expected to reduce crime.

We define violent crime as crime against another individual. This includes murder, attempted murder, manslaughter, serious wounding, rape, attempted rape, indecent assault, incest, unlawful sexual intercourse, burglary, armed robbery, robbery, attempted robbery, and other sexual advances. The Bahamas gross domestic product (GDP), and population data are publicly available (World Bank, n.d.). Violent crime data from 1963 is also publicly available from the Royal Bahamas Police Force headquarters in Nassau. Unemployment data were available from two sources; the International Labor Organization (ILO) and the Department of Statistics. The data are presented in Appendix 1.

Method

Our data allowed us to hypothesize that unemployment, population and GDP would be associated with crime. We ran single and multi-variable regressions on the data. F-tests were used to determine which factors were useful predictors (Blackwell, 2008). The main requirements to use a regression are: the predictor variables must be linearly independent, and the data is homoscedastic (the variance of the error is constant). The coefficient of determination, R^2, gives the predictive ability of the model (the percentage of the variability which is accounted for by the model.

Results

The GDP and population are independent. The labour data had a low correlation with crime, 0.053 (using ILO data) and 0.23 (using data from the Department of Statistics), to the crime rate. As the unemployment data from the Department of Statistics had a higher correlation to the crime rate, these data were used in our calculations.

Initially, we begin with linear models. We used Minitab to calculate the R^2 for every combination of one, two, and three variable linear regressions. We also ran F-tests that evaluated which variables are good predictors of the violent crime rate. The general results for the predictors of the crime rate (number of crimes/population) are below.

Linear regressions comparing the crime rate to one other variable

Crime vs unemployment (ILO data) gave R^2=.003 which implies this variable explains less than .3% of the data. Crime rate vs unemployment using government data had R^2=.053, hence it could explain 5.3% of the data (Figure 1). Running an F-test we found that the p-value is .304 which implies it is not a good predictor of the crime rate.

Using population alone we found R^2=0.542, and the p-value was close to zero. Consequently, the population variable is a useful predictor of crime rate (Figure 2). It explains 54% of the variation in the crime rate data. Using GDP, R^2=0.329 and the p-value was also near zero (Figure 3). Hence, the GDP is also a good predictor of crime rate. It explains 32.9% of the variation in the data.

Linear regressions comparing the crime rate to several variables

Including both population and GDP in the model resulted in $R^2=0.775$ with a p-value near zero, consequently adding GDP accounted for an additional 43.5% of the variation in the data being accounted for. Adding unemployment to the regression resulted in a lower R^2. This is may be caused by a reduction in the amount of information available for the model as unemployment only has 22 data points, most of which came from 1986 onwards.

Polynomial regressions

The scatterplot of GDP appeared to have a second degree variable. As a result, we examined second degree polynomials for both variables. Models with polynomial regressions with both first and second degree population and GDP terms were used. The initial regression resulted in $R^2=90.2\%$. However, this required five variables and all of them were useful predictors ($p>0.05$) of the crime rate. Using a backward elimination process, resulted in two variables: GDP^2 and $population^2$ as useful predictors ($p<0.05$) (Figure 4). This resulted in a $R^2=86.3\%$ which was the highest R^2attained with only two variables; this model using $population^2$ and GDP^2 explained 86.3% of the variation in the crime rate data.

We concluded that the square of the population and the square of GDP are useful predictors of the violent crime rate. The regression equation was:

$$\text{crime rate} = - 0.00313 - .000000000270 \times GDP^2 + 0.000000000000188 \times \text{population}^2.$$

These coefficients are rather small (because the data are large), so using normalized data (zero mean and unit variance), the equation becomes:

crime rate =0.00479- 0.00395×(GDP^2-11281671)/14629972+ 0.00534×($population^2$-58494105117)/28445081033. Note the mean of GDP^2=11281671 and $population^2$=58494105117. The standard deviation of GDP^2=14629972 and $population^2$=28445081033

Discussion

The coefficient of $population^2$ in our model is positive and so an increase in population is associated with an increase the crime rate. This indicates that as the population rises, the rate of crime rises, that is, the number of crime is disproportionate to the population. This suggests that as the population increases, government may need to invest an ever greater proportion of its resources in dealing with crime as the number of crimes increase.

However GDP^2 has a negative coefficient; hence, a higher GDP is associated with lower crime rates. Consequently, as the wealth of the nation increases, this should suppress crime. Clearly, government policies should be designed to increase the prosperity of the nation, but what these data show is that when the country cannot position itself to compete or cannot cope with external shocks, then crime would be expected to rise. Consequently, recessions should be expected to be associated with rises in crime.

Unemployment failed to be a good predictor of the crime rate. The reduced number of data points for unemployment meant that the analysis including this variable had less information from which to estimate the parameters and so would be expected to result in less precise estimates. The fact that the two sources of unemployment data disagreed with each other suggests that there may be inherent reliability issues with the data, or differences due to definitions. Analyses with more data may yet identify relationships which were not apparent the current information. Alternatively, unemployment may not be simply associated with crime, as people may live off their savings when first unemployed, so weakening the association through a time lag, or unemployment in certain industrial groups may have disproportionate consequences for crime. Such subtleties would not be evident at this global level.

GDP^2 is linear with respect to $Population^2$, the relationship is defined via $GDP^2 = .00069629 \times Population^2$, hence a change in $Population^2$ by one person would require GDP^2 to change to increase by 696 dollars. At the current population of 341,713 an increase in population of 1,000 people would require GDP to increase by 26.38 million to keep the crime rate constant.

References

Bahamas Population.(n.d.). *World Bank, World Development Indicators - Google Public Data*. Retrieved June 10, 2011, from http://www.google.com/publicdata?ds=wb-wdi&met=sp_pop_totl&idim=country:BHS&dl=en&hl=en&q=bahamas population

Black, S. (n.d.). Major Police Operation or PR Stunt?. *Bahamas B2B: Complete Guide to the Islands of The Bahamas - BahamasB2B.com*. Retrieved October 23, 2011, from http://www.bahamasb2b.com/news/2011/01/major-police-operation-or-pr-stunt-5812.html

Blackwell, M. (2008, December 3). Multiple Hypothesis Testing: The F-test. *Matt Blackwell Research*. Retrieved June 10, 2011, from http://www.mattblackwell.org/files/teaching/ftests.pdf

McElroy, J., & Roccanti, A. (2005). Crime in the Caribbean: Provisional Evidence. *Insula*, *14.2*, 32-37.

Statistics by Country (n.d.). *LABORSTA Internet (E)*. Retrieved June 10, 2011, from http://laborsta.ilo.org/STP/guest

Smith, I. (2010, July 29). Bank Robbers At Large. *The Bahama Journal*, p. 1.

Thompson, T. (2009, November 21). 18 Tourists in Shotgun Terror. *The Tribune*, p. 1.

World Bank, World Development Indicators-Google Public Data Explorer GDP (n.d.). *Google*. Retrieved June 21, 2011, from http://www.google.com/publicdata?ds=wb-wdi&met=sp_pop_totl&idim=country:BHS&dl=en&hl=en&q=bahamas%20population#ctype=l&strail=false&bcs=d&nselm=h&met_y=ny_gdp_mktp_cd&scale_y=lin&ind_y=false&rdim=country&idim=country:BHS&ifdim=country&hl=en&dl=en

Figure 1: A scatterplot depicting the limited correlation between crime rate and unemployment.

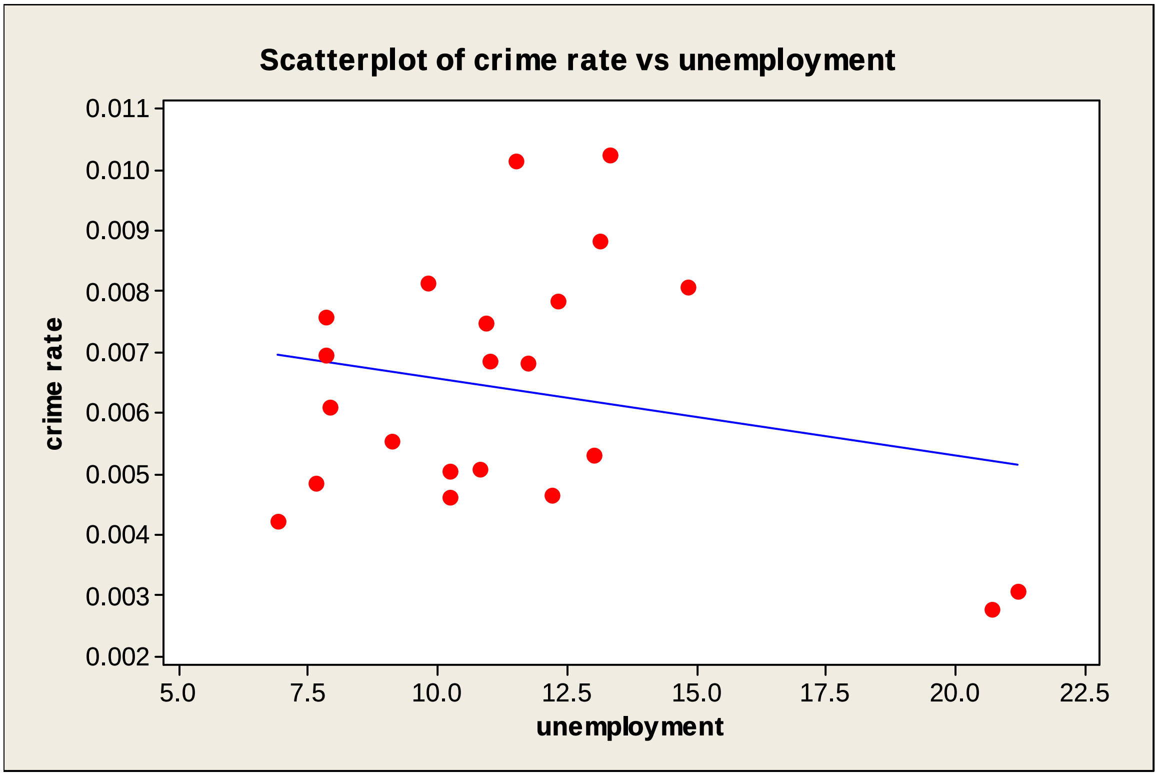


Figure2: A scatterplot depicting the correlation between crime rate and population .

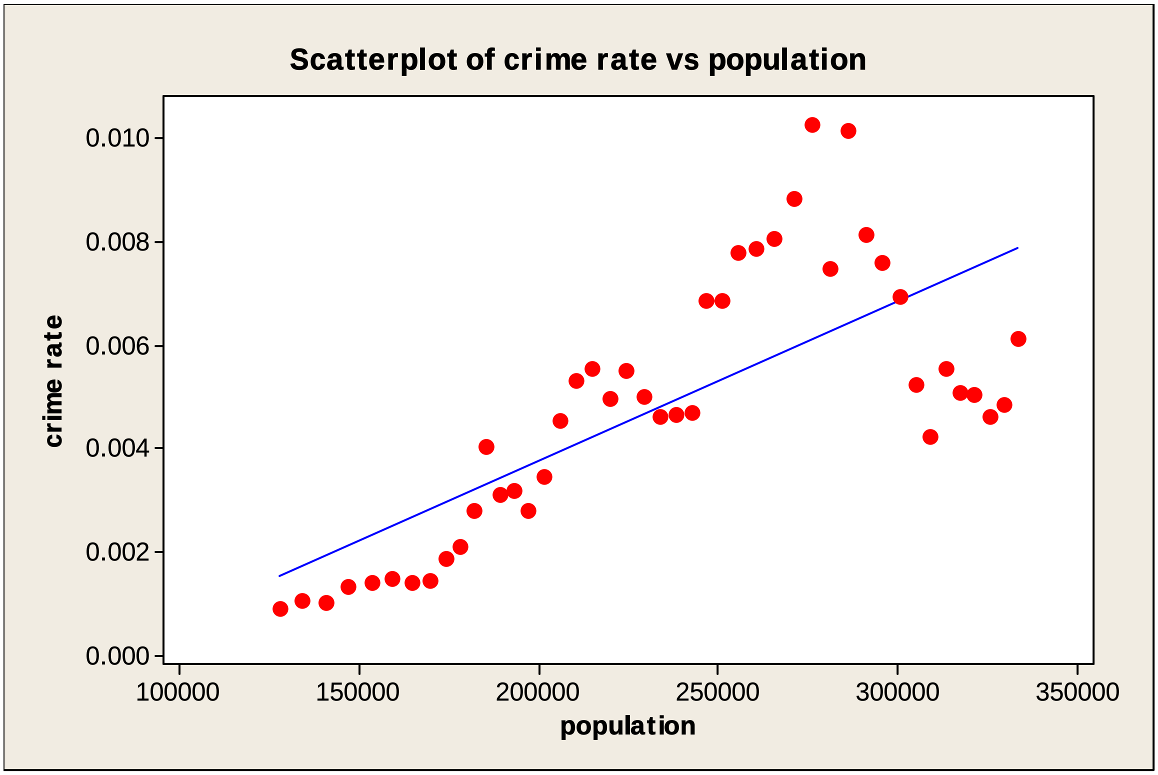


Figure 3: A scatterplot depicting the correlation between crime rate and GDP)

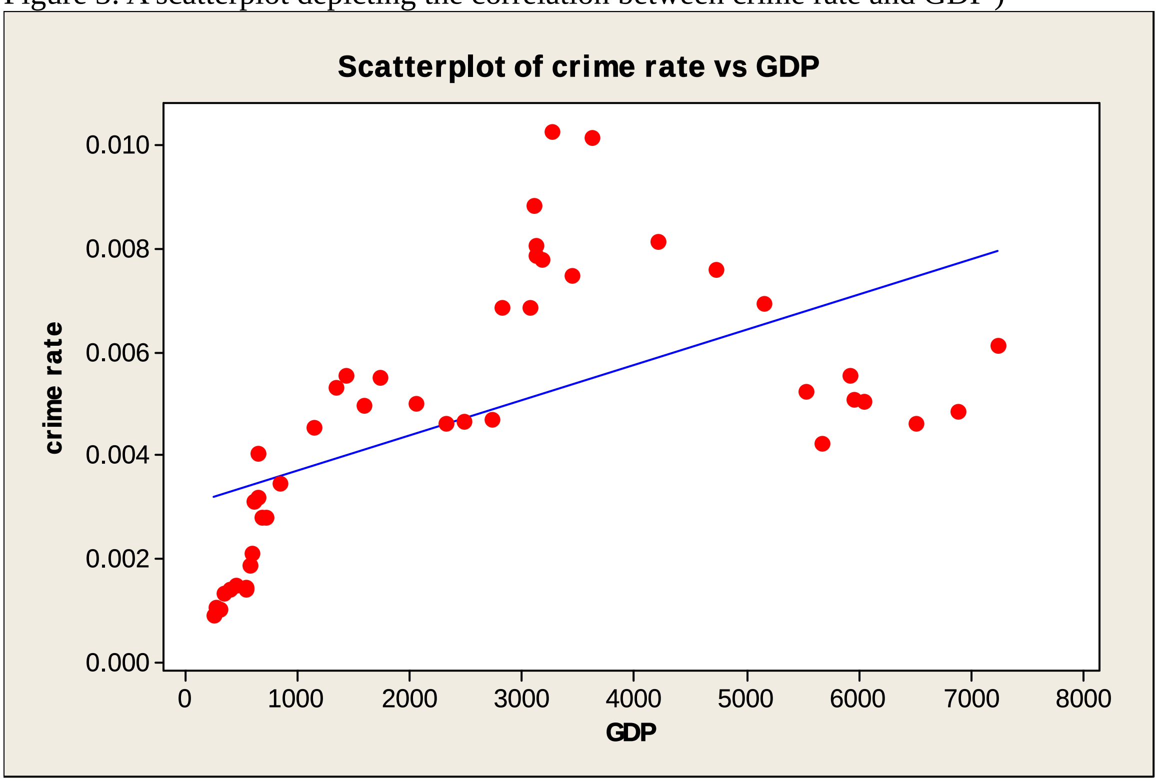
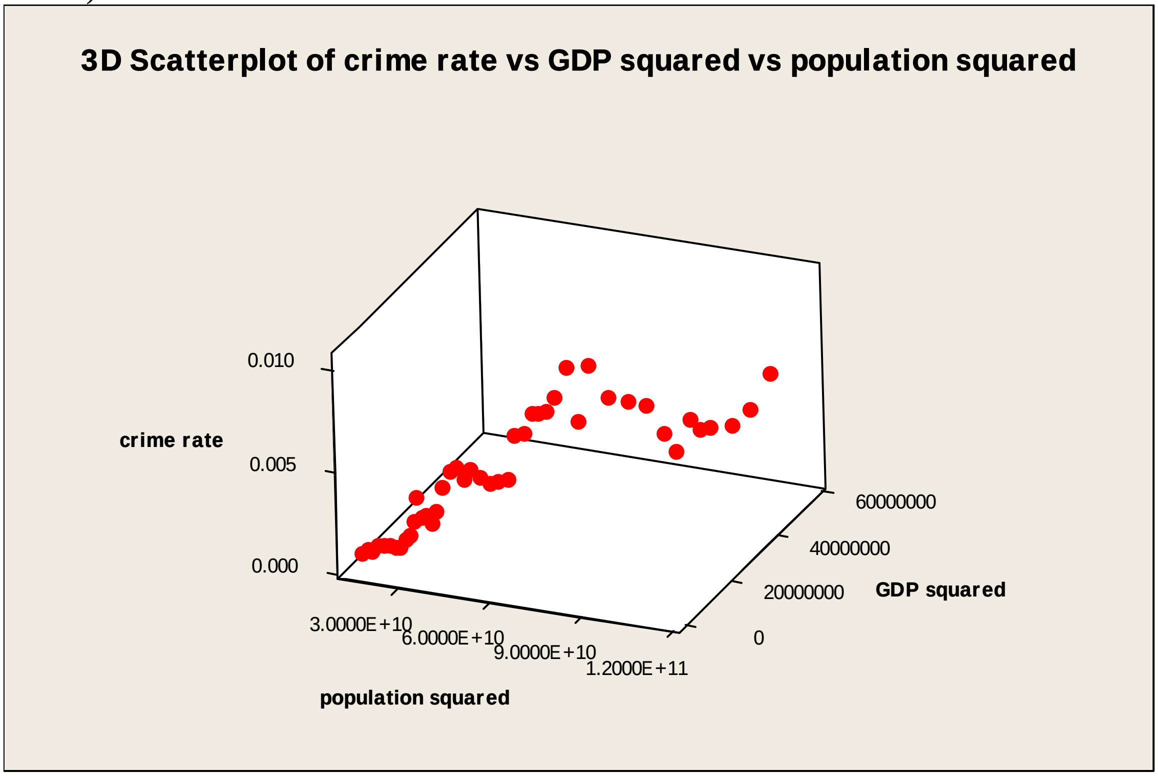


Figure 4: A scatterplot depicting a large correlation between crime rate and population^2 and GDP^2)



Appendix 1:

Data on crime, employment, population and GDP from 1960-2007

Year	Number Crimes	IOL data	Unemployment	Population	Population/crimes	Crimes/population	GDP
1960				109644			169.802
1961				115284			190.096
1962				121327			212.253
1963	115			127644	1109.947826087	0.0009009432	237.743
1964	142			134072	944.1690140845	0.0010591324	266.664
1965	140			140470	1003.3571428572	0.0009966541	300.392
1966	197			146813	745.2436548223	0.001341843	340
1967	217			153068	705.3824884793	0.0014176706	390.196
1968	233			159082	682.7553648069	0.0014646534	444.902
1969	232			164679	709.8232758621	0.0014088014	528.137
1970	243			169744	698.5349794239	0.0014315675	538.423
1971	328			174197	531.0884146341	0.0018829257	573.4
1972	376			178097	473.6622340426	0.0021112091	590.9
1973	506	6.74		181655	359.0019762846	0.0027855	670.9
1974	748			185165	247.5467914439	0.0040396403	632.4
1975	583	17.87	21.2	188851	323.9296740995	0.0030870898	596.2
1976	611			192774	315.5057283142	0.0031695146	642.1
1977	548	19.42	20.7	196889	359.2864963504	0.0027832941	713
1978	694			201181	289.886167147	0.0034496299	832.4
1979	933	11.33		205601	220.3654876742	0.0045379157	1140
1980	1114		13	210109	188.6077199282	0.0053020099	1335
1981	1192			214732	180.144295302	0.0055511056	1427
1982	1093			219482	200.8069533394	0.0049799072	1578
1983	1234			224280	181.7504051864	0.005502051	1733
1984	1145			229019	200.0165938865	0.0049995852	2041
1985	1077			233626	216.9229340761	0.0046099321	2321
1986	1107	13.5	12.2	238054	215.044263776	0.0046502054	2473
1987	1142			242338	212.2049036778	0.0047124264	2714
1988	1691	13.7	11	246593	145.8267297457	0.0068574534	2818
1989	1718	14.91	11.7	250977	146.0867287544	0.0068452488	3062
1990	1994			255603	128.1860581745	0.0078011604	3166
1991	2049	16.03	12.3	260507	127.1386041972	0.0078654316	3111
1992	2144	20	14.8	265633	123.895988806	0.0080712863	3109
1993	2391	17.95	13.1	270885	113.2936010038	0.0088266238	3092
1994	2831	19.1	13.3	276123	97.5354998234	0.0102526772	3259
1995	2106	15.59	10.9	281241	133.5427350427	0.0074882396	3429
1996	2911	16.87	11.5	286213	98.3211954655	0.010170747	3609
1997	2375	14.68	9.8	291059	122.5511578947	0.0081598576	4205

Year	Number Crimes	IOL data	Unemployment	Population	Population/crimes	Crimes/population	GDP
1998	2243	12.14	7.8	295771	131.8640213999	0.0075835697	4714
1999	2089	12.29	7.8	300354	143.778841551	0.0069551263	5150
2000	1591			304812	191.5851665619	0.0052196108	5528
2001	1310	11.33	6.9	309133	235.979389313	0.0042376582	5659
2002	1740	15.28	9.1	313319	180.0683908046	0.0055534455	5912
2003	1608	18.83	10.8	317407	197.3924129353	0.0050660508	5942
2004	1626	17.99	10.2	321453	197.6955719557	0.0050582822	6032
2005	1502	18.18	10.2	325496	216.7083888149	0.004614496	6509
2006	1594	13.83	7.6	329551	206.7446675031	0.0048368841	6876
2007	2041	14.62	7.9	333609	163.4536991671	0.0061179405	7234