Crime Index VS. Average Years of Education

Meghan Rathie, Ilinca Sipos, Avi Ahuja

Abstract:

The relationship between a country's crime index and education level is an important thing for policy makers to study because it can help make a city or country a safer place for the citizens. There are other variables that may affect crime index, but the original study began with the belief that the higher a country's education level, the lower the crime index. After research, we found that the original hypothesis was true. We added five more variables that we believed would have a significant impact on crime index to make a multiple regression model. In general, these variables behaved in the same way we had hypothesized, with a few exceptions. To try to find the best model, we ran several t-tests and f-tests to eliminate insignificant variables and see which variables had the most significant effect on crime index.

I. Introduction

A central goal which countries' have in common is the reduction of the crime rate, which would not only bring about social benefits, but can also bring economic benefits as well. Crime is a big problem, and brings a lot of social cost with it. It can economically harm countries in many ways; reducing the value of real estate, lowering tourism, and increasing tax for prisons are just some examples to show the toll which crime takes. Since crime has such a large social cost, even the lowest reductions in the crime rate can impact the economy significantly. This is why policy makers are always trying to find new and innovative solutions to reduce crime; they have looked at prison reform, bulked police forces, and many more diverse strategies. One characteristic which hasn't been studied as heavily is education, which has just started to undergo extensive quantitative and qualitative analysis.

Higher Education allows an individual to acquire new skills to make himself more marketable and useful in the work force. It gives an individual more opportunities in the market and allows for growth in many different subjects. Many people think that education only provides a private return to the individual pursing the degree, however, this is not the only benefit of higher education. Studies show that generally, education raises the productivity and earnings of an individual and those around him in the work force. Often, crime is associated with lower wealth as well, and in the cases of robbery, assault, and theft, it has usually been proved to be the case. With higher education levels, people are more likely to have the capacity to maintain a higher income than what they would be making with a lower education level. Higher average education also points to a more robust and healthy economy. Countries now have skilled workers which brings in foreign investment, as well as breeds innovations. Because of education's positive affect on the economy and social well-being, there are many reasons to believe that education will have a negative effect on the crime rate. The rational behind this statement is as follows: higher education increases the return from work which increases the opportunity cost of illegal behavior. A long with this, education can have profound affects on an individual alone, such as making said individual more patient, risk averse, and logical leading an individual to more sensible decisions.

We hypothesize that countries with higher levels of education, higher average ages, a lower Gini coefficient, higher GDP per capita, lower population density, and a lower male-to-female ratio will have lower crime rates. Through rigorous statistical analysis, specifically both, single and multiple linear regression, we hope to come to conclusions which support our hypothesis. The output of these tests (granted they agree with our hypothesis and policy makers pursue education reform), can have

profound affects on society as a whole by giving people and policy makers alike a basis on which to act on when it come to policies and decisions.

II. Literature Review

We have found three journals relating the impact of high education levels to the decrease in crime rates. Below are also validations and explanations for why we chose to regress the particular variables that we chose.

Lance Lochner and Enrico Moretti (2004) study prison inmates to find the effect of education on crime. They compile data on incantation by using the FBI database. Since, they collect data partitioned by the type of crime Lochner and Moretti are also able to find the type of crime that education affected the most. The researchers also use two other data sets namely, state-level data on arrests from the Uniform Crime Reports, and self-report data on crime and incarnation from the National Longitudinal Survey of Youth. Lochner and Moretti were able to reach the same conclusion from all three of these data-sets, that schooling reduces criminal activity. Using both, the OLS and IV method they were able to justify this decision quantitatively. Lochner and Moretti both support the notion that the positive returns gained from schooling on society is greater than the positive returns gained privately estimating that a 1% increase in the high school graduation rate in the United States would save the country $1.4 billion on crime.

Machin, Marie, and Vujic (2010) study how education reduces the crime rate. They start off with the notion that education has a negative affect on the crime rate. These researchers used data from three sources to get all the relevant information. Sources used by the researchers are as follows: Offenders Index Database (England and Wales), Census, and the British Crime Survey. Results from this study are consistent with the hypothesis that higher education levels can have a substantial impact on the crime rate. To quantitatively prove this hypothesis the researchers used a regressions discontinuity approach. Machin, Marie, and Vujic suggest improving education amongst potential and present offenders is an approach which policy makers should pursue in the future.

Randi Hjalmarrson and Lance Lochner (2012) study the impact of education on crime using international evidence to support their hypothesis that improving education can have a significant impact on the crime rate. Hjalmarrson and Lochner use FBI's Uniform Crime Reports which has data spanning from 1960-1990. The researchers are analyzing state-level male arrests based on age and criminal offense. The strategy which Hjalmarrson and Lochner are using is to measure how a state's

change in compulsory schooling age leads to a change in the higher education rate and subsequently the crime rate. Hjalmarrson and Lochner reach 6 main conclusions. They are as follows: increasing educational attainment levels yields substantial social benefit, policies which increase high school graduation rate will have the most affect on the crime rate, policies targeting at risk groups will have more affect on the crime rate, a higher education attainment rate will decrease both, property and violent crimes, if an individual has higher wages their opportunity cost of violent and property crimes rise, and lastly educational attainment does not have to increase to reduce crimes, better schooling has more of an effect.

Jesse Brush (2007) studies the effect that income inequality has on crime. He starts with the assumption that the national income inequality level has a larger effect on crime than the income inequality within smaller regions. Brush uses data from the US Census Office to compile about income inequality and crime. He found that there was a significant positive correlation between the Gini coefficient and crime rates. Brush then added other variables such as median income, population density, the percent of people between ages 18-24, unemployment rate and percent of people in different ethnic groups to his model. From the new variables added, Brush's main take away point was that high-income areas have the lowest crime. In the end, Bursch states that there are many factors affecting crime rates, with income inequality being one of those factors.

Farrington (1986) studies the relationship between age and crime. Typically, crime tends to peak during teenage years, and from there decreases gradually. Farrington finds that although the age-crime curve varies by individual, the aggregate curve peaks around age 17. Farrington also studies the types of crimes and their frequency by age groups, finding that more severe crimes happen between the ages of 25 – 30. This shows how a younger average age can be connected to higher crime rates.

Darrell Steffensmeier and Emilie Allan (1996) study the relationship between gender and crime. It is universally recognized that females commit fewer crimes than males. Men commit more crime than women in all sectors of crime other than prostitution. Areas with higher male rates of crime also have higher female rates of crime. They also suggest like other factors associated with gender, such as schooling and up-bringing, can play a part in the different crime rates between males and females.

Dennis Roncek (1975) studies the effect of urban density on crime. For his data, Roncek uses he number of offenses that were reported to the police as the dependent variable. Population density was the independent variable. He finds that overcrowded areas tend to have higher crime rates, but there

are social variables that come with living in population dense areas that also play a role in higher crime levels. Unfortunately, these social and cultural variables that are inevitably present are very difficult to quantify in a manner that allows them to be included in a regression model.

Daniel Lederman, Norman Loayza and Ana María Menéndez (2002) study how economic development changes crime rates. They start the study by noting the GDP per capita has been an important determinant in violent crime rates in the past, and hypothesize that an increasing GDP per capita will lower crime rates. For their data, Lederman, Loayza, and Menéndez use The World Bank to get World Development Indicators of GDP. They also studied social capital and crime, so in turn had to note the how factors like GDP, income inequality, and quality of police effect social capital and in turn effect crime. They found that GDP per capita was a significant variable in determining crime rates.

III. Data

Crime and education are two variables that are hypothesized to have a negative correlation. As education level increases, by country, crime levels are expected to decrease. The dependent variable, crime, changes based on average education levels in each country. As the average education level goes up, crime is expected to go down because people's behavior is expected to change in a positive way and therefore reducing violence.

The 2013 crime index was found at a database called Numbeo. There was a list of 118 countries listed on the site for the 2013 crime index, 73 of which matched up with the listed average education levels by country that were listed on the UNESCO website. Numbeo is a database that is used by credible business magazines such as *The Economist* and *Forbes*, as well as in *The New York Times* and *BBC.* The UNESCO data was compiled based on the reporting of each country. Since both are impartial sources, the crime index values and average education levels are expected to accurately reflect how education levels affect crime rates by country.

Variable	Data Source
Y (Crime index)	Numbeo
Average Education	UNESCO
GINI coefficient	CIA
Male to Female Ratio	CIA
Average Age	CIA
Population Density	Worldbank
GDP Per Captia	Knomea Database

The data for average age came from the CIA World Factbook, which provides information about 267 countries. The Factbook has information about a country's history, people, government, economy, energy, communications, transportation, military, and transnational issues. Out of all the countries listed that had information for average age, all 73 we had from the simple regression model matched. The CIA gave average female and average male age separately, but for out data we used the average age of the total population because this is more applicable information when looking into crime levels.

The data for the male to female ration also came from the CIA World Factbook. The Factbook split up the ratio into many different age sectors, but for our data we used the male to female ratio for the total population. Out of the countries listed, 73 matched the data we had for the other variables in the model. The sex ratio is expected to have a positive correlation with crime index.

The data for the GINI coefficient was also taken from the CIA World Factbook and the countries that were listed matched the 73 countries we had information for. The GINI coefficient for each country was not taken in the same year, but because it was taken around the same time period, the statistics should not be skewed. Countries with more income inequality are expected to have higher crime rates.

The population density data was taken from The World Bank. The World Bank is an organization that aims to end world poverty, and their website offers information about hundreds of different countries to the public. The listed countries under the population density information matched the 73 countries that we had been using.

The GDP per capita was taken from the Knoema Database. Knoema compiles information from various databases such as The World Bank, IMF, and UNESCO. Again, all 73 of the countries used in our previous data matched with the countries listed on this data source.

The table below states what the variables we used are comprised of and what exactly they are measuring, as well as what format.

Label	Variable description	Type of Measure	Type
crimeindex	The index of crime by country based on a survey given to the people in the country, which is based off the national standards of crime	0-100, measure made off of a country's crime relative to other countries	Dependent
avgeducation	Average years of education in a country including primary, secondary, tertiary, and above	Years	Independent

AvgAge	The average age of the people in the country	Years	Independent
GINIcoefficient	The GINI coefficient of a country giving measure of income inequality	Measure of statistical dispersion, 0-100	Independent
lngdpcapita	The natural log of the GDP per capita in the country	Natural log of dollars	Independent
lnpopdens	The natural log of the population density in the country	Natural log of population	Independent
MaletoFemale	The ratio of males to females populating a country	Ratio	Independent

The countries in which we gathered data for the above come from very diverse backgrounds. We have countries ranging from rich to poor, undeveloped to developed, etc. The list of the countries we used are stated in the appendix.

In our simple regression model, we only regressed crime rates against the average years of education, while in our multiple regression models, we used a variety of variables. The descriptive statistics of all of the variables we used are shown in the table below.

Variable	Observations	Mean	SD	Minimum	Maximum
Average Education	64	15.08528	2.413816	7.79477	20.43272
Average Age	64	34.87188	7.156337	21.4	46.5
GINI Coefficient	60	36.63167	8.963778	23.7	57.7
ln(GDP per capita)	64	9.843967	0.8128454	7.720905	11.31132
ln(Population density)	64	4.451066	1.249557	1.102033	8.831309
Male to Female Ratio	64	0.979375	0.0792399	0.85	1.41
Crime Index	64	42.28203	12.9536	13.11	84.87

Out of the 64 observed countries, the average crime index was 42.3 (out of 100). Crime index is measure of the overall crime in a given country. A crime level below 20 is considered very low, crime levels between 20-40 are considered low, crime indices between 40-60 indicate moderate crime levels, and anything above 60 is considered very high. An average crime level of 42.9 shows that overall, among all the countries surveyed, crime levels are between moderate and low. An average education level of 15.027 shows that on average, a country's citizens spend around 15 years in schools, which means many people probably attend some sort of college or university. The standard deviation for crime is fairly large,

meaning that there is a lot of variability between crime indices at certain education levels. The standard deviation for education is moderately large. A deviation of 2.444 years can be a big change in the amount of schooling the average person receives.

The analyzed data is only reliable if the Gauss Markov assumptions are met for our data. Below are the validations for each of the assumptions based on our data.

1. Linearity in Parameters

   Our model can be written in the form $Y = \beta0 + \beta1X1 + \beta2X2 + \beta3X3 + \beta4X4 + \beta kXk + u$, meaning it is linear in parameters.

2. Random Sampling of variables

   The data included information of each variable from countries all over the world that had complete information lining up with our variables, with no calculated method of picking and choosing. The model for the random sample is $Yi = \beta0 + \beta1Xi1 + \beta2Xi2 + \beta3Xi3 + \beta4Xi4 + \beta kXik + ui$

3. No Perfect Collinearity

   As the table below shows, there is no perfect collinearity between any of the independent variables. We purposely chose variables though we would not think are highly correlated.

Variable	Average Education	GINI Coefficient	Male to Female	Average Age	LN(GDP Per Capita)	LN(Population Density)
Average Education	1.0	N/A	N/A	N/A	N/A	N/A
GINI Coefficient	-0.3733	1.0	N/A	N/A	N/A	N/A
Male to Female	-0.2602	0.1446	1.0	N/A	N/A	N/A
Average Age	0.0970	0.0229	-0.0381	1.0	N/A	N/A
Ln(GDP per Capita)	0.7370	-0.2179	-0.1274	0.1613	1.0	
Ln(Population Density)	-0.2709	0.0768	0.0966	-0.0103	-0.0905	1.0

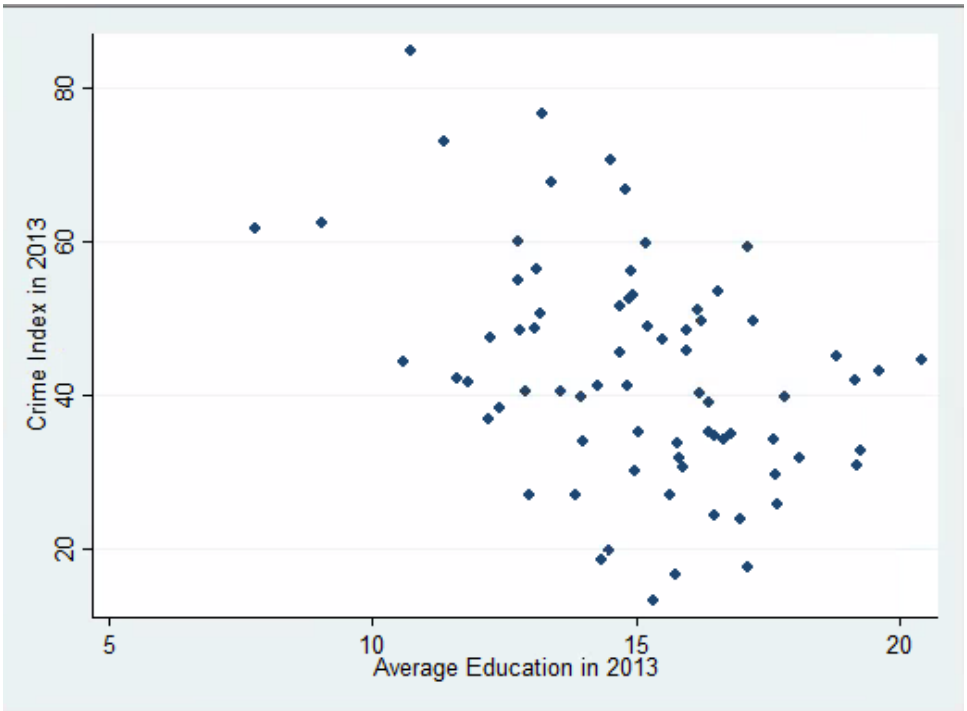
4. Zero Conditional Mean

   The expected value of the error term is zero, such that $E(u) = 0$.

5. Homoskedasticity

Our last assumption was that the variance in the error term is the same for all of the chosen independent variables.

III. Results




```
. regress crimeindexin2013 avgeducationin2013
```

Source	SS	df	MS	Number of obs	=	73
				F(1, 71)	=	14.17
Model	2569.2101	1	2569.2101	Prob > F	=	0.0003
Residual	12870.0424	71	181.268203	R-squared	=	0.1664
				Adj R-squared	=	0.1547
Total	15439.2525	72	214.434063	Root MSE	=	13.464

crimeindexin2013	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
avgeducation~2013	-2.444031	.6491838	-3.76	0.000	-3.738466	-1.149595
_cons	79.63489	9.881723	8.06	0.000	59.9313	99.33849

In our simple regression model, the coefficient of our x variable, average education, is notable. As shown in the regression table above, the coefficient is -2.444, meaning for every unitary increase in the average years of education, it is estimated that the crime index for the country will drop 2.44 points. As seen in the scatterplot of the simple regression model, the data is not tightly packed, which is also

reflected in the R squared value (0.1664). This value is pretty small compared to what we would want in more ideal regression model. The hope for the multiple regression model is for it to capture many of the other major factors of crime, making the data more compact and bringing up the R squared value.

Multiple Regression:

Model 1:

crimeindex = 79.63 -2.44(*avgeduation*)

Model 2:

crimeindex = **30.88** - **1.24***(avgeduation)* + **.0.203***(AvgAge)* - 2.39*(lngdpcapita)* + **0.52***(GINIcoefficient)* + **46.12***(MaleToFemale)* - **3.73***(lnPopulationDensity)*

Model 3:

crimeindex = 79.58 -**1.58**(*avgeduation*) +**0.54**(*GINIcoefficient*) - **1.64**(Lngdpcapita) - **3.70**(Lnpopdens)

Independent Variables:	Model 1	Model 2	Model 3
avgeducation	-2.44*** (-3.76)	-1.24 (-1.20)	-1.58 (-1.55)
GINI		0.517*** (2.83)	.54*** (2.93)
Ln(gdpcapita)		-2.39 (-0.84)	-1.64 (-0.56)
Ln(popdens)		-3.73*** (-2.92)	-3.70*** (-2.88)
AvgAge		.203 (0.94)	
MaletoFemale		46.12 (1.52)	
Intercept	79.63*** (8.06)	30.88 (0.82)	79.58*** (3.49)
No. of Observed	73	60	60
R-squared	0.1664	0.325	0.276

***significant at 10% , **significant at 5%, ***significant at 1%, T-statistic in parentheses**

By looking at the table above one can observe the coefficient and t-statistic values for all the variables across the three regression models. In the first model, we found average education to have a coefficient of –2.44 and it was significant all the way through to the 1% level. To work towards finding our optimal regression model, we decided to use all of our variables in our regression model 2. From this we were able to find that only two of the variables were significant by performing the t-test. We decided

to perform F-Tests to filter out variables based on whether they were jointly significant. When performing an F-Test, we found avgeducation and ln(popdens) to output a value of 4.54 with the parameters F(2,60). This is greater than the critical value of 3.15 meaning that avgeducation and ln(popdens) are jointly significant. Due to this we decided to keep both these variables in the model. We found that AvgAge and MaletoFemale were not jointly significant with each other or any of the other variables in our model, and decided to remove these variables from the model.

This leads us to finding our optimal model, model 3. In this model we see that when the t-test is performed, GINIcoefficient, ln(popdens), and the intercept are all statically significant at the 1% level. We also found an R-squared value of 0.276.

V.Conclusions

The main purpose of this study was to find which macroeconomic variables had the largest impact on crime index. Our simple regression model had the trends that we expected. We found that as average years of education increased, the average crime index decreased. This was not enough information to really see what affects the crime index, and since the main purpose behind this project was to put together information to potentially help policy makers implement plans for safer communities, we needed to add more variables to form a multiple regression model.

We began our multiple regression model by compiling a list of factors we believed would affect the crime index. We hypothesized that a higher Gini coefficient, male to female ratio, and population density would lead to higher crime rates. We believed that as average age and GDP increased, the crime rates would decrease. Other than population density having a negative effect on crime rates, all other variables had the effect we expected. It was odd that as population density increased, the crime index decreased because we expected places to get less safe as more people lived in an area. We also concluded that average age and the male to female ratios were not significant factors when it comes to crime rates. Cultural differences play a big role in crime rates by country. Although we recognized this, there was no way of finding a quantifiable measure of cultural differences, so we could not include that as one of our variables.

Our optimal model consisted of the following variables: average education, GINI coefficient, and the natural log of both, GDP per capita and population density. It makes sense for these variables to be included in the optimal model, as they all play a big part in how much crime takes place in a country.

Further studies should include more variables so researchers are able to derive more evidence to back how average education inversely affects the crime rate of a country. Additionally, it would be beneficial to include more countries in the studies and maybe divide countries into different bins. This means separating countries by region, GDP, gun policy, and a multitude of other classifications.

References

Brush, Jessee. "Does income inequality lead to more crime? A comparison of cross-sectional and time-series analyses of United States counties," Economics Letters, Volume 96, Issue 2, August 2007, Pages 264-268

Farrington, David P.. “Age and Crime”. *Crime and Justice* 7 (1986): 189–250. Web.

Lederman, Daniel, Norman Loayza, and Ana María Menéndez. “Violent Crime: Does Social Capital Matter?”. *Economic Development and Cultural Change* 50.3 (2002): 509–539. Web.

Lochner Lance and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, 94(1):155-189.

Randi Hjalmarsson & Lance Lochner,2012.The Impact of Education on Crime: International Evidence CES for DICE Report,Ifo Institute for Economic Research at the University of Munich, vol. 10(2), pages 49-55, 08.

Roncek, D. W. (1975). Density and crime A methodological critique. *The American Behavioral Scientist (Pre-1986), 18*(6), 843. Retrieved from http://prx.library.gatech.edu/login?url=http://search.proquest.com/docview/194650846?accountid=11107

Steffensmeier, Darrell, and Emilie Allan. “Gender and Crime: Toward a Gendered Theory of Female Offending”. *Annual Review of Sociology* 22 (1996): 459–487. Web...

Stephen Machin & Olivier Marie & Sunčica Vujić, 2010."The crime reducing effect of education,"LSE Research Online Documents on Economics 28727, London School of Economics and Political Science, LSE Library.

APPENDIX

List of all countries:

Albania	Bulgaria	Ecuador	Guatemala	Italy	Lithuania	Pakistan	Serbia
Argentina	Chile	Egypt	Honduras	Japan	Malaysia	Panama	Slovenia
Australia	China	Estonia	Hungary	Kazakhstan	Malta	Philippines	Sweden
Austria	Hong Kong	Finland	India	Kuwait	Mauritius	Poland	Thailand
Azerbaijan	Costa Rica	France	Indonesia	Kyrgyzstan	Mexico	Portugal	Tunisia
Belarus	Cyprus	Georgia	Iran	Laos	Nepal	Republic of Moldova	Turkey
Belgium	Czech Republic	Germany	Ireland	Latvia	New Zealand	Russian Federation	Ukraine
Brunei	Denmark	Greece	Israel	Lebanon	Norway	Saudi Arabia	USA

```
. regress crimeindexin2013 avgeducationin2013
```

Source	SS	df	MS	Number of obs	=	73
				F(1, 71)	=	14.17
Model	2569.2101	1	2569.2101	Prob > F	=	0.0003
Residual	12870.0424	71	181.268203	R-squared	=	0.1664
				Adj R-squared	=	0.1547
Total	15439.2525	72	214.434063	Root MSE	=	13.464

crimeindexin2013	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
avgeducation~2013	-2.444031	.6491838	-3.76	0.000	-3.738466	-1.149595
_cons	79.63489	9.881723	8.06	0.000	59.9313	99.33849

```
. regress crimeindex avgeducation AvgAge GINIcoefficient lngdpcapita lnpopdens MaletoFemale
```

Source	SS	df	MS	Number of obs	=	60
				F(6, 53)	=	5.73
Model	4616.51719	6	769.419532	Prob > F	=	0.0001
Residual	7111.18903	53	134.173378	R-squared	=	0.3936
				Adj R-squared	=	0.3250
Total	11727.7062	59	198.774682	Root MSE	=	11.583

crimeindex	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
avgeducation	-1.238939	1.032726	-1.20	0.236	-3.310326	.8324473
AvgAge	.2025664	.2143862	0.94	0.349	-.2274379	.6325706
GINIcoefficient	.5173017	.1827215	2.83	0.007	.1508089	.8837945
lngdpcapita	-2.39122	2.929001	-0.82	0.418	-8.266052	3.483612
lnpopdens	-3.731575	1.276121	-2.92	0.005	-6.29115	-1.172001
MaletoFemale	46.12015	30.37092	1.52	0.135	-14.79619	107.0365
_cons	30.87918	37.59681	0.82	0.415	-44.53048	106.2888

```
. correlate avgeducation GINIcoefficient MaletoFemale AvgAge lngdpcapita lnpopdens
(obs=60)
```

	avgedu~n	GINIco~t	Maleto~e	AvgAge	lngdpc~a	lnpopd~s
avgeducation	1.0000					
GINIcoeffi~t	-0.3733	1.0000				
MaletoFemale	-0.2602	0.1446	1.0000			
AvgAge	0.0970	0.0229	-0.0381	1.0000		
lngdpcapita	0.7370	-0.2179	-0.1274	0.1613	1.0000	
lnpopdens	-0.2709	0.0768	0.0966	-0.0103	-0.0905	1.0000

Source	SS	df	MS	Number of obs	=	60
				F(4, 55)	=	7.67
Model	4198.23548	4	1049.55887	Prob > F	=	0.0001
Residual	7529.47074	55	136.899468	R-squared	=	0.3580
				Adj R-squared	=	0.3113
Total	11727.7062	59	198.774682	Root MSE	=	11.7

crimeindex	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
avgeducation	-1.580442	1.019943	-1.55	0.127	-3.624453	.4635686
GINIcoefficient	.5387891	.1841165	2.93	0.005	.1698113	.9077669
lngdpcapita	-1.640902	2.921713	-0.56	0.577	-7.496145	4.214341
lnpopdens	-3.709211	1.288892	-2.88	0.006	-6.292208	-1.126215
_cons	79.57637	22.79318	3.49	0.001	33.89781	125.2549