The Effect of High School Drop Out Rates on Crime Rates

Jesus Barrera Mejia, Myles Smith

ECON 3112 001

5/2/2023

The Data

Research Question

- *What is the effect of high school dropout rates on the crime rate?*
- Our question was what is the effect of high school dropout rates on the crime rates? The crimes we included were violent crimes & property crimes. The total crimes include assault, battery, and rape, in addition to property crimes like burglaries and motor-related crimes.

The Variables

Total_crime_rate_per_100k	Violent & property crime per 100,000 people
Drop_out_rate	Public High school dropout rate
Illicit_drug_use_per_100k_1000	Drug use per 100,000 people, in thousands (Not Including Marijuana)
spending _per_pupil_1000	Spending per pupil for public schools, in thousands of dollars
Ginix100	Gini index percentage points, income inequality
Medical	1 if state has medical marijuana legislation
Recreational	1 if state has recreational marijuana legislation

*Omitted Variable: Illegal -> State has illegal marijuana laws

- The was also another variable, GDP per capita but it had an extreme effect on all the other variables causing many of them to lose their significance and change the results significantly in addition to having some correlation with other variables, so it was omitted in the regression.

Using Describe

```
 Observations:         2,049
    Variables:             9                  17 Apr 2023 13:31
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
Total_crim~100K double  %10.0g                Total_crime_rate_per_100K
Drop_out_rate   double  %10.0g                Drop_out_rate
Illicit_dr~100k double  %10.0g                Illicit_drug_use_per_100k
Spending_per_~l double  %10.0g                Spending_per_pupil
Ginix100        double  %10.0g                Ginix100
Ginix100_Sqr    double  %10.0g                Ginix100_Sqr
Medical         byte    %10.0g                Medical
Recreational    byte    %10.0g                Recreational
Illegal         byte    %10.0g                Illegal
-------------------------------------------------------------------------------
```

- When running the describe function all the variables are formatted correctly and none of them are strings.

Using Summarize

```
. summarize
```

Variable	Obs	Mean	Std. dev.	Min	Max
Total_c~100K	153	2775.514	814.2781	1421.7	6008.8
Drop_out_r~e	153	5.407962	1.445645	2.654035	10.33383
Illicit~100k	153	2837.737	438.1529	2077.559	4561.348
Spending_p~l	153	12506.69	3659.472	6953	24048.21
Ginix100	153	46.71784	2.064626	40.81	54.2
Ginix100_Sqr	153	2186.792	194.8311	1665.456	2937.64
Medical	153	.3986928	.4912373	0	1
Recreational	153	.2352941	.4255756	0	1
Illegal	153	.3660131	.4832951	0	1

- Using the summarize function shows us that the data in the data set looks good, although the massive gaps between the minimum and maximum for dropout rates, spending per pupil, and crime rate may cause slight concern.
- For the dummy variables only 23.5% of the states in the 153 observations had recreational marijuana compared to 39.87% that are medical and 36.60% that have legislation making it illegal. This is meaningful because we have a slightly limited idea of how recreational marijuana laws affect crime rate, compared to medical and illegal legislation.

Correlation Matrix

	Tot~100K	Drop_o~e	Ill~100k	Spendi~l	Ginix100	Ginix1~r	Medical	Recrea~l	Illegal
Total_c~100K	1.0000								
Drop_out_r~e	0.4643	1.0000							
Illicit~100k	0.2515	0.0366	1.0000						
Spending_p~l	-0.2714	-0.5166	0.3062	1.0000					
Ginix100	0.2016	0.1765	0.2107	0.1466	1.0000				
Ginix100_Sqr	0.2066	0.1742	0.2154	0.1510	0.9993	1.0000			
Medical	-0.1261	-0.0042	-0.0061	0.2386	0.0777	0.0773	1.0000		
Recreational	0.2363	-0.0414	0.4592	0.2013	0.0242	0.0282	-0.4517	1.0000	
Illegal	-0.0800	0.0407	-0.3981	-0.4198	-0.1003	-0.1035	-0.6187	-0.4215	1.0000

- Here the interesting, but expected correlation is between dropout rates and spending per pupil, there is a strong negative correlation meaning as spending per pupil increases dropout rate decreases, which makes logical sense.
- Another notable strong correlation is between the state marijuana legislation and illicit drug use, there seems to be a moderately strong positive correlation between states with recreationally legal marijuana and illicit drug use, illicit drug use per 100k does not include marijuana, meaning states with recreationally legal marijuana may have increased drug use.

Regression

Total_crime_rate_per_100k	Violent & property crime per 100,000 people
Drop_out_rate	Public High school dropout rate
Illicit_drug_use_per_100k_1000	Drug use per 100,000 people, in thousands (Not including marijuana)
spending _per_pupil_1000	Spending per pupil for public schools, in thousands of dollars
Ginix100	Gini index, income inequality
Medical	1 if state has medical marijuana legislation
Recreational	1 if state has recreational marijuana legislation

*Omitted Variable: Illegal -> State has illegal marijuana laws

Source	SS	df	MS		
Model	35604496.5	7	5086356.65	Prob > F	0.0000
Residual	65178927	145	449509.841	R-squared	0.3533
Total	100783424	152	663048.839	Root MSE	670.45

Number of obs	=	153
F(7, 145)	=	11.32
Prob > F	=	0.0000
R-squared	=	0.3533
Adj R-squared	=	0.3221
Root MSE	=	670.45

Total_crime_rate_per_100K	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Drop_out_rate	184.9942	47.77977	3.87	0.000	90.55943	279.429
Illicit_drug_use_per_100k_1000	313.8473	151.4408	2.07	0.040	14.53079	613.1639
spending_per_pupil_1000	-52.75221	20.82892	-2.53	0.012	-93.91974	-11.58469
Ginix100	-1440.634	721.5598	-2.00	0.048	-2866.768	-14.50053
Ginix100_Sqr	15.84706	7.654334	2.07	0.040	.7185781	30.97554
Medical	30.22761	138.5755	0.22	0.828	-243.6613	304.1166
Recreational	400.9929	172.9706	2.32	0.022	59.12349	742.8623
_cons	34086.92	17037.53	2.00	0.047	412.9306	67760.9

- R Squared
  - 32.21% of the variation in violent crime rates is explained by high school dropout rates, illicit drug use, education spending per pupil, income inequality, & the state's marijuana legislation, Ceteris Paribus.
- Dropout Rate
  - As the high school dropout rate increases by 1 percentage point, the crime rate increase by 185 crimes per 100,000 people, ceteris paribus.

    - Drop out rate is statistically significant in this model, the t > 1.96, the probability the null hypothesis is true is 0%, and 0 is not included in the confidence interval.

- Illicit Drug Use
    - As illicit drug use increases by 1000 users per 100,000 people, Violent crimes increase by 314 crimes per 100,000 people, Ceteris Paribus
    - Illicit drug use is statistically significant with a t > 1.96 and the probability of the null hypothesis being true of 4%, a little close but good nonetheless. In addition, 0 is not in the confidence interval

- Spending Per Pupil
    - As spending per pupil increases by $1000, Violent crimes increase by 53 crimes per 100,000 people, ceteris paribus
    - Spending per pupil is statistically significant with a t > 1.96 and the probability of the null hypothesis being true of 1.2%. In addition, 0 is not in the confidence interval

- Gini Index
    - As the Gini index increases, violent crimes decrease at a decreasing rate until the Gini index is equal to 45.45, as the Gini index continues to increase violent crime increases at an increasing rate, ceteris paribus
    - The Gini Index, or income inequality, is statistically significant. Both t's are greater than 1.96. The regular Gini index has a probability of the null hypothesis being true

- Medical
    - States with medical marijuana laws have a crime rate per 100,00 people of 30 more, compared to illegal states, ceteris paribus.
    - Medical is not statistically significant, we know because the t-value is less than 1.96 and the probability the null hypothesis is true is 82.8 %, in addition, 0 falls in the confidence interval.

- Recreational
    - States with recreational marijuana laws have a crime rate per 100,000 people of 400.99 more than states with illegal marijuana legislation, ceteris paribus.
    - Recreational marijuana legislation does have statistical significance with a t-value of 2.32 greater than 1.96, and a probability of a true null hypothesis of 2.2 %, in addition to not having 0 in the confidence interval.

Post-Regression

The key takeaway from formulating a research question, brainstorming variables, gathering data, & running the regression was how incredible of a tool linear regression is. The process was helpful in learning one of the key tool's economists use in their trade and how it allows researchers to provide decision-makers with quantifiable answers to questions, however, during the process some key pitfalls were discovered. One of the biggest challenges was finding data & ensuring the quality of the data, there are a plethora of different sources provided by the internet, however, it is important to ensure the source is trustworthy and understand how their datasets were created, predominantly data from third-party sources. In addition, when running the regression, the data in the data set must be properly formatted and clean from any errors, because this can throw off results and cause errors when running a regression in Stata or other software.

The process provided practice for the process of conducting econometric research. The process begins with a research question and can continue until the researcher is satisfied with the model which can prove to be a trap, we find ourselves delaying the regression and constantly attempting to find new variables to add to the model. This process clarified how variables interact and how adding on a variable can have a massive impact or no impact at all on the regression. Most importantly the process provided us an opportunity to implement the tools we learned in class, including interpretations of the variables, and creating new kinds of variables using existing variables, most notably the squared terms.

During the regression analysis, the most interesting effect I found was that of the Gini index. When left alone, it is not statistically significant to the crime rate, however, once the term is squared it becomes significant and provides an interesting interpretation. The model only explains 32% of the variation in crime rates, however, the Gini index provides an interesting result. It states that as income inequality increases, the crime rate decreases at a decreasing

rate. When interpreting this it means that there is an acceptable level of income inequality, however, once income inequality begins to increase towards its highest level, crime rates increase at an increasing rate. To illustrate this point think about how crime would be high if everyone was poor, which would be a world with no income inequality, compared to a world where there is a large gap in income inequality, both could lead to high crime. This indicates the best scenario is when it's in the middle. However, this result may change with more data and could require more research, but nonetheless, it provides an interesting foundation for future research.

The current model is a good foundation, but there are more variables that could help increase the adjusted R squared, for example, unemployment, homelessness, and welfare usage. In addition, the current model only has observations from 2016 to 2018, including other years could yield different results. Otherwise, the model is great and was successful in providing an answer to the primary research question of the effect of dropout rates on crime rates.