

## MULTIPLE OLS REGRESSION

### RESEARCH QUESTION ONE:

#### Predicting State Rates of Robbery per 100K

We know that robbery rates vary significantly from state-to-state in the United States. In any given state, we also know that robbery rates vary across smaller geographical units like counties and even neighborhoods. What factors are related to murder rates at the aggregate state level? Let's use State level data from 2012 to find out. There are several theories of violence that help us understand state-to-state fluctuations of violent crime including robbery. These include social disorganization theory that postulates structural factors such as economic deprivation and residential mobility affect a community's ability to informally control its residents. One measure of economic deprivation is the percent of individuals that live below the SSA level of poverty. A measure of residential mobility is the percent of the population in a community that has moved in the past 5 years. Robberies are also more likely to occur in urban locations, so another factor that would affect robbery rates in states is the percent of the state's population that resides in rural locations. We would predict that the higher rate of rurality, the lower the rate of robbery in each state. And finally, regional location has typically been related to violent offending. Generally, states located in the South have higher rates of violence than states in the nonSouth.

For this exercise, then, we will use state rates of robbery per 100K as the dependent variable along with 4 independent variables: 1) percent living in poverty, 2) percent who have moved in past 5 years, 3) percent residing in rural areas, and, 4) regional location.

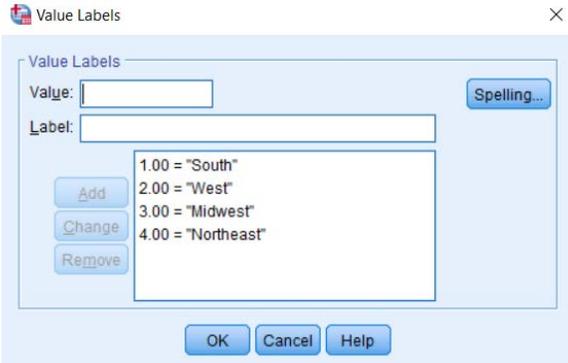
In this data set, let's run descriptives for each of our variables:

**ANALYZE /Descriptive Statistics /Descriptives**

#### Descriptive Statistics

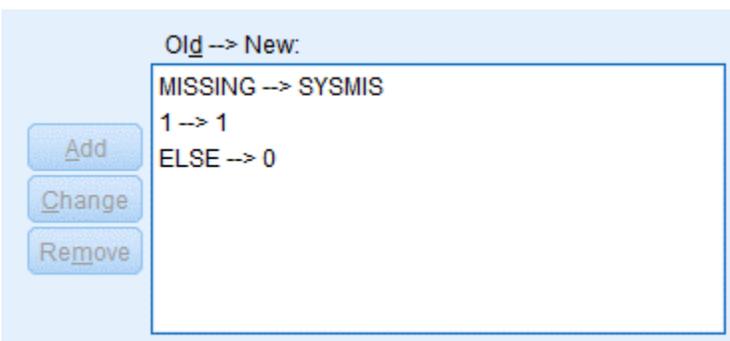
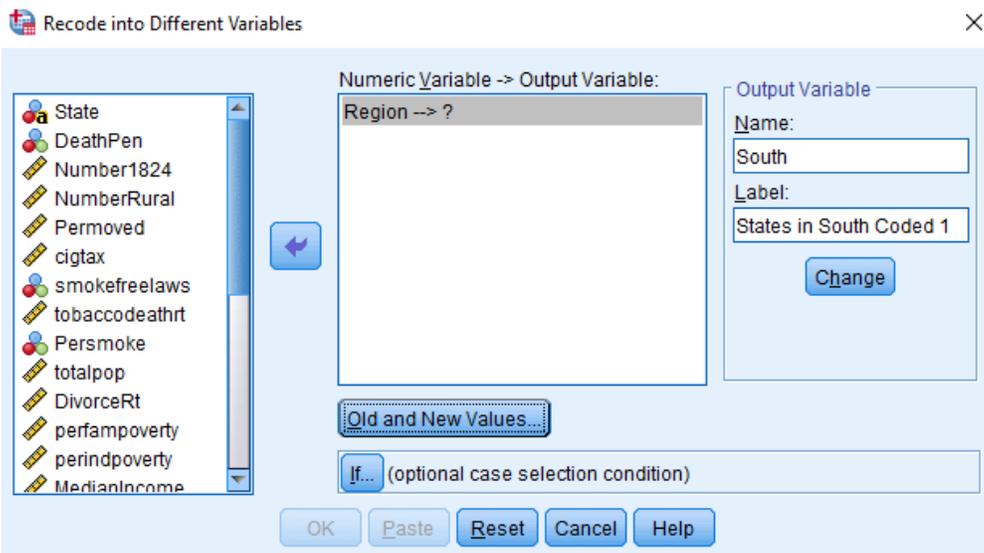
	N	Minimum	Maximum	Mean	Std. Deviation
RobberyRt Robbery Rate per 100K	50	14.30	260.70	104.5840	57.73535
Permoved Percent in different house in 2010	50	9.20	21.50	15.2220	2.43531
PercentRural Percent of Pop Living in Rural Areas	50	5.05	60.06	26.0740	14.24530
Region Four regional categories	50	1.00	4.00	2.2800	1.10730
perfampoverty Percent of Families below poverty	50	5.50	17.30	9.8740	2.75309
Valid N (listwise)	50				

One thing to notice right way is that all the variables are interval/ratio level except for region. It ranges from 1 to 4. If we take a look at the value labels of this variable, we see that states in the South are coded 1 and all other regions are coded from 2 to 4.



Because every type of regression including OLS regression is not be able to interpret a nominal/categorical level variable like this, we need to recode it into a dichotomy coded “0” and “1.” *Note: When I recode a variable into a dichotomy, I always name it what the value of 1 is going to equal.* For example, in this case, I am going to create a variable called “South” that is coded “1” for states that reside in the South and “0” for states that reside in the NonSouth. Using this coding convention makes it easier to interpret your results – for example, if I had named this variable region, I would probably have to keep referring back to the value labels or codebook to see how it was coded.

## TRANSFORM /Recode into Different Variables



The easiest way to do this would be to code 1 equal to 1 and all other values equal to 0, remembering to code all missing values set to missing just in case! One thing SPSS requires is that you press the **“Change”** button before the program will actually recode the variable. This is just a safety mechanism to ensure that researchers

actually want to change a variable, especially when they are not recoding into a new variable. After you hit “Change” you can then press “OK” and you will see the newly created variable at the bottom of the variable list in the Variable View screen of SPSS.

Remember to run a frequency distribution of both the original regional variable and the new variable called South to ensure the recode has been done correctly:

**ANALYZE /Descriptive Statistics /Frequencies**

**Region Four regional categories**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 South	17	33.3	33.3	33.3
	2.00 West	13	25.5	25.5	58.8
	3.00 Midwest	12	23.5	23.5	82.4
	4.00 Northeast	9	17.6	17.6	100.0
	Total	51	100.0	100.0	

**South States in South Coded 1**

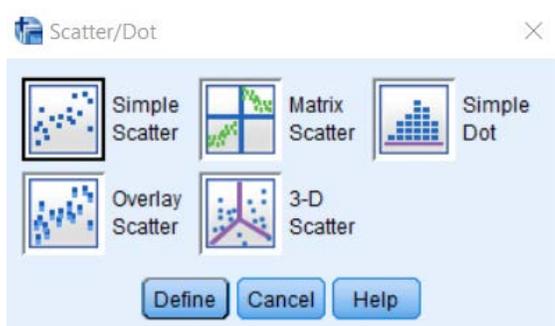
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	34	66.7	66.7	66.7
	1.00	17	33.3	33.3	100.0
	Total	51	100.0	100.0	

From the output above, we can see that the 17 cases, or 33.3% of the state distribution that was coded South in the original regional variable is now coded “1” in the South variable.

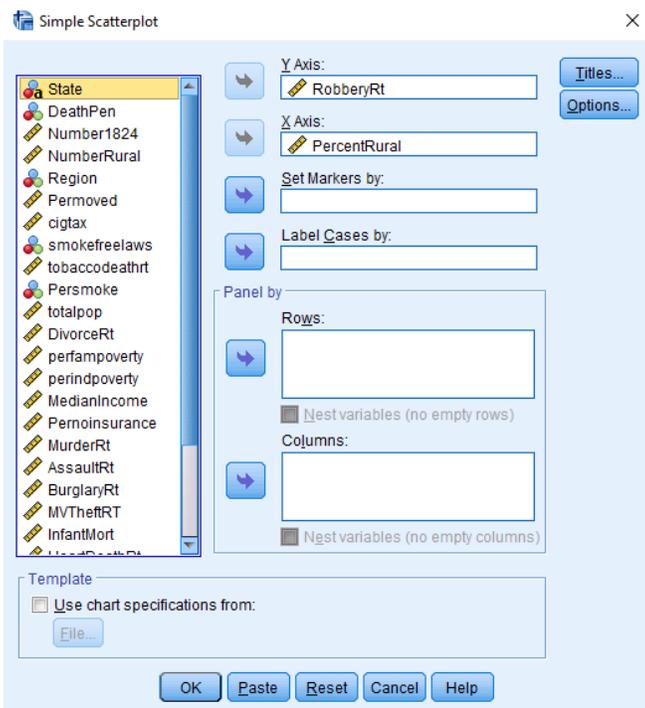
Since these data are state level, there is a small enough sample size to make examining the bivariate scatterplot of each IV and DV combination important. Let’s first examine the relationship between the percent urban and robbery rates for states. We can do this Using the Graphs command:

**GRAPHS /Legacy Dialog /Scatter-Dot**

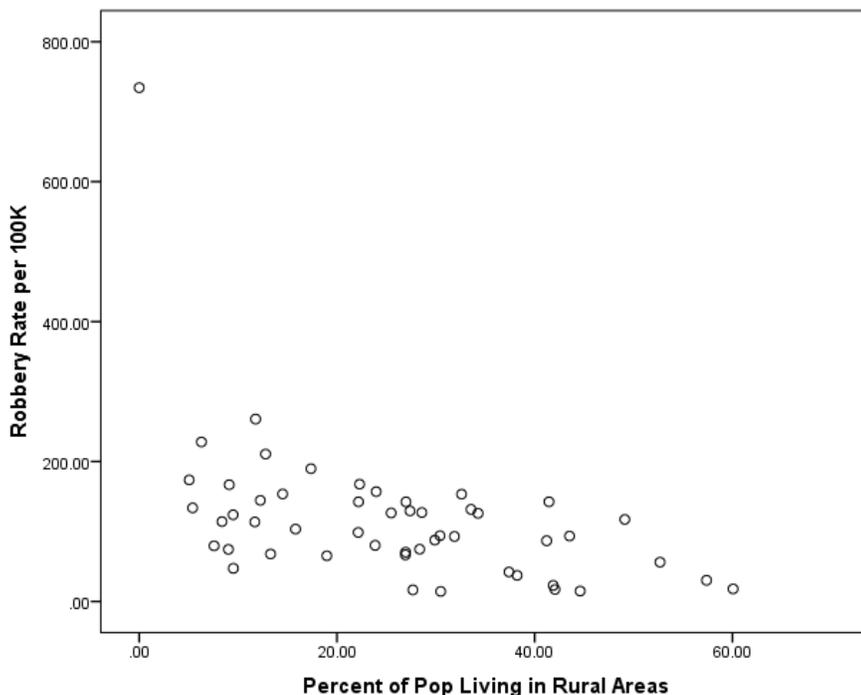
For now, let’s just do a simple scatterplot:



Click “Define” and place the **robbery rate on the Y axis** and the **percent rural variable on the X axis**:

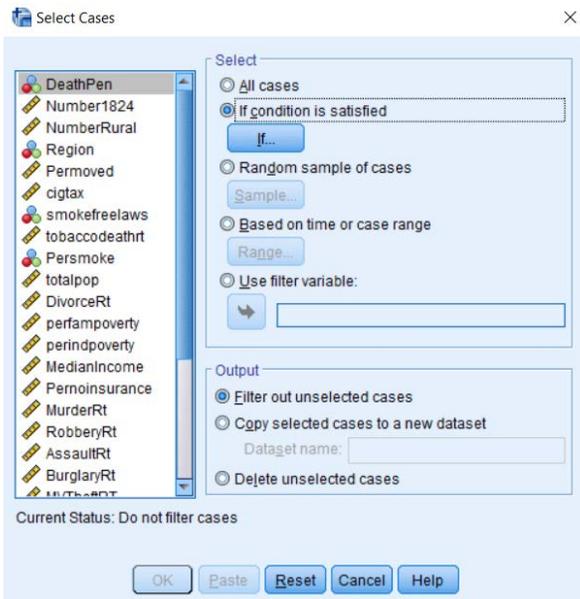


This produces the following scatterplot:



One thing is clear from this scatterplot, and that is a very significant bivariate outlier. If we examine our data, we will see that this is the District of Columbia, which has a 0 percent of its population residing in rural areas, and a very high robbery rate. There is really no hard and fast rule for determining what constitutes a bivariate outlier, but a data point like this that falls significantly away from the rest of the data could easily be justified as an outlier. *What should we do about these?* In this case, it is easiest to go in call this data point missing for the multivariate analysis. One easy way to do this in SPSS is through the “select cases” option, which is under the “Data” Icon.

## DATA /Select Cases



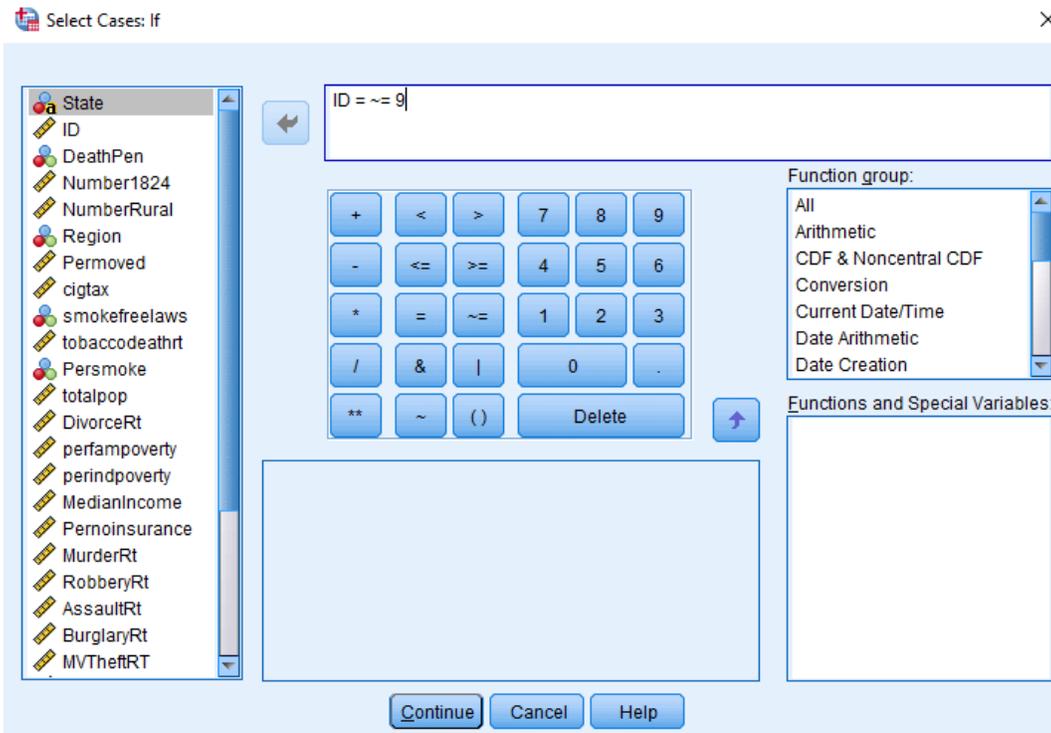
This gives us several options from which to select. In this case, we are simply going to select the criteria that selects cases “If condition is satisfied” and that condition is if the case is equal to DC. There are several ways we can do this. This data set includes an ID number associated with each case. If we examine the data in the “Data View” window, we can see that the ID associated with DC is 9:

The screenshot shows the 'Data View' window in SPSS. The title bar reads '\*State.2012.b..sav [DataSet2] - IBM SPSS Statistics Data Editor'. The menu bar includes 'File', 'Edit', 'View', 'Data', 'Transform', 'Analyze', and 'Direct Marketin'. Below the menu bar is a toolbar with icons for file operations and data manipulation. The main area shows a table with two columns: 'State' and 'ID'. The 'State' column lists states from Alabama to Georgia, and the 'ID' column lists corresponding numerical values from 1.00 to 11.00. A red arrow points to the row where 'State' is 'District of Columbia' and 'ID' is 9.00.

	State	ID
1	Alabama	1.00
2	Alaska	2.00
3	Arizona	3.00
4	Arkansas	4.00
5	California	5.00
6	Colorado	6.00
7	Connecticut	7.00
8	Delaware	8.00
9	District of Columbia	9.00
10	Florida	10.00
11	Georgia	11.00

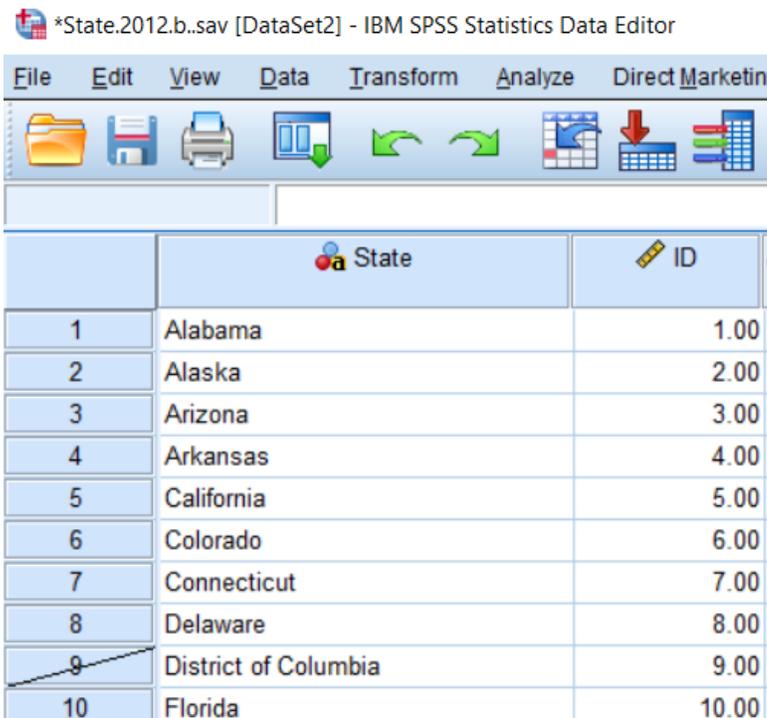
Therefore, our goal is to select all cases in which the ID variable is not equal to 9. We can easily do this under the “Select Cases” Option of the Data dialogue box:

## DATA /Select Cases /If Condition is Satisfied

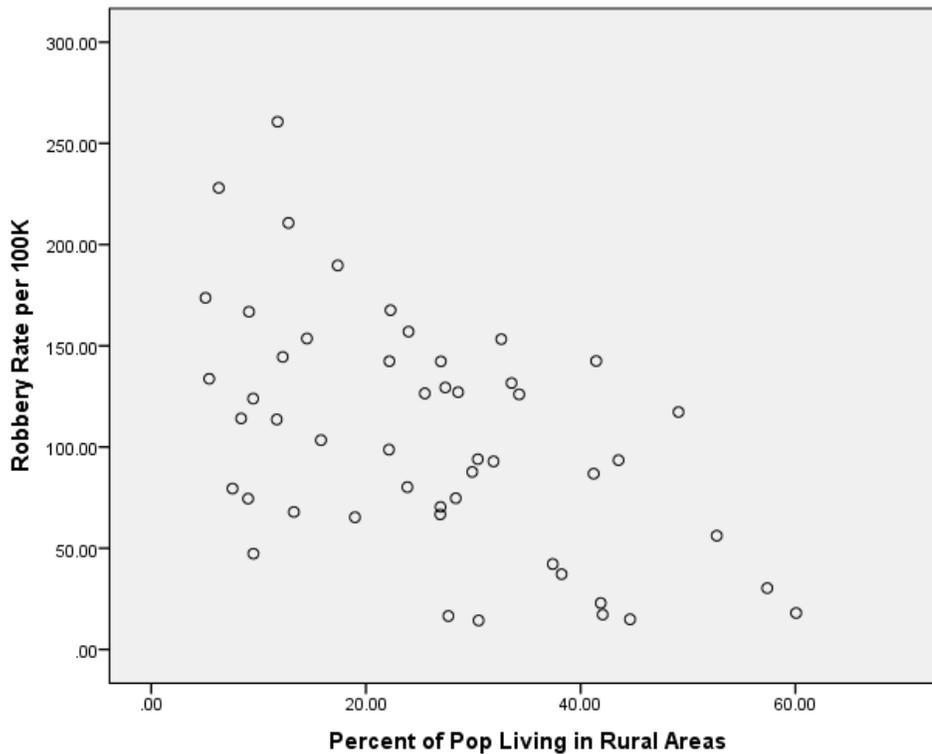


If you have several bivariate outliers, it may be best to simply eliminate them from your data by setting them as “missing” in the in the “**Variable View**” menu. For now, we will simply set DC as missing for the analyses to follow using this Select If function.

SPSS actually provides a visual for cases that are excluded by this command in the “Data View” window. As you can see below, if you toggle to the Data View, you can see that DC has a hash mark through it, indicating that it is no longer included in the data analysis:



Now let's take another look at that scatterplot between rurality and robbery rates in states:



We see a very different picture here. I won't go through each IV and DV relationship as I did in the bivariate webinar, but you should inspect all your bivariate relationships using the Graph Function, just like you should inspect all univariate distributions to determine that there is not severe skewness or outliers as well.

For now, we are going to move on to other preliminary bivariate analysis using the correlation function. We need to do a preliminary investigation to ensure that multicollinearity is not an issue:

**ANALYZE /Correlate /Bivariate**

**Correlations**

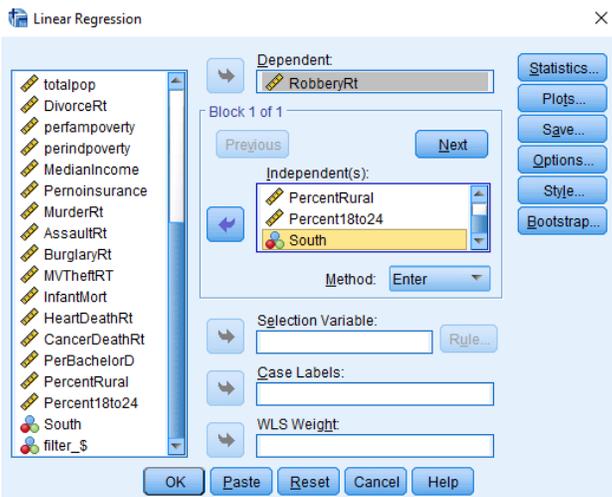
		RobberyRt Robbery Rate per 100K	Permoved Percent in different house in 2010	PercentRural Percent of Pop Living in Rural Areas	South States in South Coded 1	perfampoverty Percent of Families below poverty
RobberyRt Robbery Rate per 100K	Pearson Correlation	1	-.091	-.533**	.320*	.256
	Sig. (2-tailed)		.529	.000	.023	.073
	N	50	50	50	50	50
Permoved Percent in different house in 2010	Pearson Correlation	-.091	1	-.017	.022	.117
	Sig. (2-tailed)	.529		.907	.878	.417
	N	50	50	50	50	50
PercentRural Percent of Pop Living in Rural Areas	Pearson Correlation	-.533**	-.017	1	.218	.255
	Sig. (2-tailed)	.000	.907		.129	.073
	N	50	50	50	50	50
South States in South Coded 1	Pearson Correlation	.320*	.022	.218	1	.576**
	Sig. (2-tailed)	.023	.878	.129		.000
	N	50	50	50	50	50
perfampoverty Percent of Families below poverty	Pearson Correlation	.256	.117	.255	.576**	1
	Sig. (2-tailed)	.073	.417	.073	.000	
	N	50	50	50	50	50

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Any correlation that is high between any combination of the IV's may be a sign that multicollinearity will be a problem. *What is high?* The general rule is above  $r = .75$ . If you find that, it is generally a good idea to create an index or omit one of these IVs from the multiple regression model. I don't see any problems here so we can move on to the multiple regression. We don't see high correlations between the IV's here so we can move on to our multiple regression.

**ANALYZE /Regression /Linear**



## Regression

### Variables Entered/Removed<sup>a</sup>

Model	Variables Entered	Variables Removed	Method
1	Percent of Families below poverty, Percent in different house in 2010, Percent of Pop Living in Rural Areas, States in South Coded 1 <sup>b</sup>	.	Enter

a. Dependent Variable: RobberyRt Robbery Rate per 100K

b. All requested variables entered.

### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.734 <sup>a</sup>	.539	.498	40.90996

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	88022.037	4	22005.509	13.148	.000 <sup>b</sup>
	Residual	75313.130	45	1673.625		
	Total	163335.167	49			

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	159.894	42.155		3.793	.000
	Permoved Percent in different house in 2010	-3.321	2.423	-.140	-1.371	.177
	PercentRural Percent of Pop Living in Rural Areas	-2.719	.426	-.671	-6.376	.000
	South States in South Coded 1	39.283	15.254	.321	2.575	.013
	perfam poverty Percent of Families below poverty	5.425	2.656	.259	2.043	.047

a. Dependent Variable: RobberyRt Robbery Rate per 100K

Let's first examine the model fit statistics of R and R<sup>2</sup>. Unlike these same statistics for the bivariate case, these don't tell us anything about any individual relationships, they simply tell us the relationship between all IVs in the model and the DV.

According to the multiple correlation coefficient, combined, all the IVs in this model are strongly related to the robbery rate per 100K in states. According to the adjusted R<sup>2</sup>, all of the IVs together explain almost 50% of the variation in robbery rates. According to the Model Summary F test, the model explains a significant amount of variation in robbery rates.

**The multiple regression equation according to the Coefficients output is:**

$$Y_{RobberyRt} = 159.89 + -3.32(x_{\%moved}) + -2.72(x_{\%rural}) + 39.28(x_{South}) + 5.42(x_{\%poor})$$

Looking at the significance for each individual slope coefficient, we see that that there are only three variables that significantly affect robbery rates in states:

- For every one-unit increase in rural population in states, robbery rates per 100K decrease 2.7 units, even after controlling for the other IVs. In general, then, states with higher levels of rural population tend to have lower rates of robbery.
- Compared to states in the NonSouth, there is a 39.28 unit increase in robbery rates for states that reside in the South, even after controlling for the other IVs. In general, then, states in the South have higher rates of robbery than states in the NonSouth.
- For every one unit increase in families living below the poverty level in states, there is a 5.42 unit increase in robbery rates, even after controlling for the other IVs. In general, then, states with higher rates of poverty tend to have higher rates of robbery.

*How do we assess relative strength between those variables that are significant?* Well, some statisticians will have you look at the standardized coefficients (Beta) in the output. Unlike the unstandardized regression slope coefficients, these are standardized like the correlation coefficients. However, these can sometimes lead us astray. Therefore, I simply rely on the size of the t value. Remember that a t will be negative if the relationship between x and y is negative so the sign does not imply anything about the strength of the relationship. Of those coefficients that are significant, generally meeting an alpha/significance level of .05 or lower, the highest t value indicates the strongest relationship. You could also say that the t value that achieves the greatest significance (lower risk of being wrong) is strongest, but sometimes you have more than one that attains a significance of .000 and since SPSS only displays 3 decimal places, it would be difficult to know which is lower unless you click on it to display all the values beyond the third decimal.

According to the t values, rurality has the most influence on robbery rates in states. A great way to display the effect of key variables on a dependent variable to an audience is to predict values of the DV at different levels of an important IV while holding all other variables constant at some level, typically their means, or modes in the case of categorical variables.

For example, lets predict the robbery rate for a state in the NonSouth that has an average rate of families living in poverty (9.9), an average mobility rate (15.2), and has a very low percentage of rural population (5):

$$Y_{RobberyRt} = 159.89 + -3.32(15.2) + -2.72 (5) + 39.28(0) + 5.42(9.9)$$

$$Y_{RobberyRt} = 159.89 + -50.46 + -13.6 + 0 + 53.65$$

$$Y_{RobberyRt} = \mathbf{149.48}$$

Now let's compare that to a state with the same mobility and poverty rates in the NonSouth, but that has a much higher percentage of rural population (50):

$$Y_{RobberyRt} = 159.89 + -3.32(15.2) + -2.72 (50) + 39.28(0) + 5.42(9.9)$$

$$Y_{RobberyRt} = 159.89 + -50.46 + -136 + 0 + 53.65$$

$$Y_{RobberyRt} = \mathbf{27.08}$$

Our robbery rate is 5 times lower!

## RESEARCH QUESTION TWO:

What are the factors that affect the sentence length received by a sample of homicide defendants convicted of murder? Using a sample of data from 75 of the largest Metropolitan areas in the U.S., the dependent variable we will examine is prison sentence length (in days) received.

We will examine the effects of several Independent Variables including demographic characteristics of the defendant (age, sex, race), characteristics of the murder (number of victims, relationship to victim), type of adjudication (whether the conviction was a plea bargain or a jury trial), and criminal history of the defendant.

Let's first examine the univariate distribution of each variable.

**ANALYZE /Descriptive Statistics /Descriptives**

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
PrisonTime Incarceration term in days	1008	2	73000	16080.47	14641.641
DefAge Age in yr a/o offn	1295	13.83	87.30	29.2304	10.63385
whitedefendant	1314	.00	1.00	.3661	.48191
knownvictim	1313	.00	1.00	.4775	.49969
numbervictims number of victims	1311	1	5	1.07	.332
jurytrial	1314	.00	1.00	.7146	.45177
maledefendant sex of defendant	1319	0	1	.90	.304
priorcon prior convictions	1319	.00	1.00	.3306	.47059
Valid N (listwise)	977				

We see from the output that we have a total of 977 valid homicide defendant cases. Prison sentence age defendant's age are both interval/ratio level variables, race and gender of defendant, type of trial, and victim/offender relationship are all dichotomous variables that are appropriately coded 0 and 1. The only variable that we need to recode is number of victims. The descriptives tell us that it ranges from 1 to 5 with a mean of 1.07 so we know that it is probably very skewed. Let's take a closer look at this variable with a frequency distribution.

**ANALYZE /Descriptive Statistics /Frequencies**

numbervictims number of victims					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1233	93.5	94.1	94.1
	2	63	4.8	4.8	98.9
	3	12	.9	.9	99.8
	4	1	.1	.1	99.8
	5	2	.2	.2	100.0
	Total	1311	99.4	100.0	
Missing	System	8	.6		
Total		1319	100.0		

From this, we see that the vast majority of homicides included one victim. This is an extremely skewed distribution and because it deviates so much from a normal/symmetrical distribution, it is best to recode this into a dichotomous variable coded “0” for those homicides that included one victim and “1” for all those that included 2 or more victims.

We can do this under **Transform**. *Note: I always use the option of recoding into a different variable – never recode the original variable or you will forever lose it!*

## **TRANSFORM     /Recode into Different Variables**

Using my convention of naming dichotomous variables what “1” is coded, I will call this variable “multiplevictims,” and label it Two or more victims. The other convention I use when recoding is to ALWAYS code all system-or user missing values as system missing in the first command in the Old->New dictionary.

Again, the easiest way to recode this variable with the fewest steps would be to recode 1 equal to 1 and all other values equal to 0.

Press “**Change**” and Press **OK**, and the new variable will then be added to the data set and it will be listed at the end of variable list and can be seen in the “**Variable View**.”

Again, remember to run a frequency distribution of the newly created variable to ensure that it has been recoded correctly. That is, the new variable called “multiplevictims” should have about 94.1% of the distribution coded 0 and 5.9% of the distribution coded 1. Let’s see:

## **ANALYZE   /Descriptive Statistics        /Frequencies**

### **Multiplevictims Two or more victims**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1233	93.5	94.1	94.1
	1.00	78	5.9	5.9	100.0
	Total	1311	99.4	100.0	
Missing	System	8	.6		
Total		1319	100.0		

And that is exactly what the frequency distribution shows, so we can rest assured that our recode was correct.

I am going to forgo examining the scatterplots of each IV and DV combination for this example, but you should always do that. I will inspect a bivariate correlation matrix to see the relationships between the IV's and the DV and also inspect the correlations between all of the IVs to ensure that there is not a relationship that is highlight correlated, which you will remember, indicates that multicollinearity is a problem.

## ANALYZE /Correlate /Bivariate

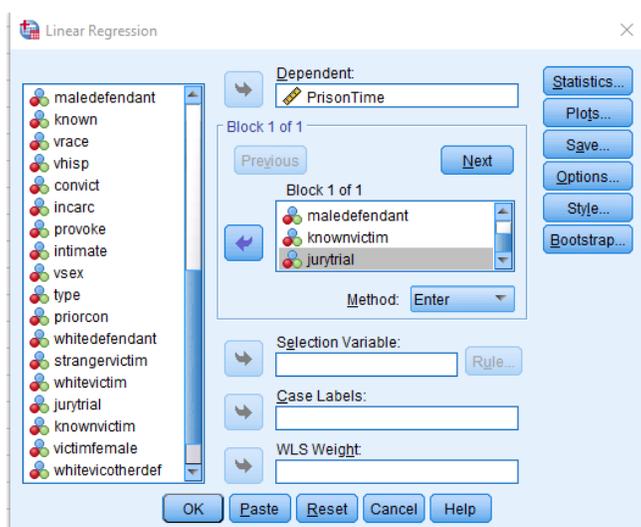
		Correlations						
		DefAge Age in yr a/o offn	whitedefendant	maledefendant sex of defendant	knownvictim	jurytrial	priorcon prior convictions	PrisonTime Incarceration term in days
DefAge Age in yr a/o offn	Pearson Correlation	1	.039	-.091**	-.049	-.058*	.148**	-.012
	Sig. (2-tailed)		.163	.001	.077	.037	.000	.710
	N	1295	1291	1295	1289	1290	1295	989
whitedefendant	Pearson Correlation	.039	1	.014	.005	.161**	-.068*	.051
	Sig. (2-tailed)	.163		.601	.851	.000	.014	.103
	N	1291	1314	1314	1308	1309	1314	1003
maledefendant sex of defendant	Pearson Correlation	-.091**	.014	1	-.093**	.136**	.164**	.097**
	Sig. (2-tailed)	.001	.601		.001	.000	.000	.002
	N	1295	1314	1319	1313	1314	1319	1008
knownvictim	Pearson Correlation	-.049	.005	-.093**	1	.030	-.017	.067*
	Sig. (2-tailed)	.077	.851	.001		.273	.538	.034
	N	1289	1308	1313	1313	1308	1313	1005
jurytrial	Pearson Correlation	-.058*	.161**	.136**	.030	1	.063*	.273**
	Sig. (2-tailed)	.037	.000	.000	.273		.022	.000
	N	1290	1309	1314	1308	1314	1314	1008
priorcon prior convictions	Pearson Correlation	.148**	-.068*	.164**	-.017	.063*	1	.066*
	Sig. (2-tailed)	.000	.014	.000	.538	.022		.037
	N	1295	1314	1319	1313	1314	1319	1008
PrisonTime Incarceration term in days	Pearson Correlation	-.012	.051	.097**	.067*	.273**	.066*	1
	Sig. (2-tailed)	.710	.103	.002	.034	.000	.037	
	N	989	1003	1008	1005	1008	1008	1008

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

I don't see any problems here so we can move on to the multiple regression.

## ANALYZE /Regression /Linear



## Regression

### Variables Entered/Removed<sup>a</sup>

Model	Variables Entered	Variables Removed	Method
1	jurytrial, DefAge Age in yr a/o offn, known known or unknown def & vict, priorcon prior convictions, whitedefendant, maledefendant sex of defendant <sup>b</sup>		Enter

### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.292 <sup>a</sup>	.085	.080	14020.121

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	17837204970.000	6	2972867496.000	15.124	.000 <sup>b</sup>
	Residual	191649695200.000	975	196563789.900		
	Total	209486900200.000	981			

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4999.126	2309.708		2.164	.031
	DefAge Age in yr a/o offn	-5.937	44.397	-.004	-.134	.894
	whitedefendant	97.066	945.912	.003	.103	.918
	maledefendant sex of defendant	3291.555	1687.954	.062	1.950	.051
	known known or unknown def & vict	1977.131	902.505	.067	2.191	.029
	priorcon prior convictions	1440.869	963.166	.047	1.496	.135
	jurytrial	8873.224	1079.124	.260	8.223	.000

a. Dependent Variable: PrisonTime Incarceration term in days

**Multiple regression equation:**

$$Y_{\text{sentence}} = 4999.12 + -5.9(x_{\text{defage}}) + 97.1 (x_{\text{whitedef}}) + 3291.5(x_{\text{maledef}}) + 1977.1(x_{\text{knowvic}}) + 1440.8(x_{\text{prior}}) + 8873.2(x_{\text{jurytrial}})$$

*Note: A word about individual-level data before we move on. There is always much more measurement error and variation with individual level data compared to aggregate level data. As a result, the correlation coefficients and  $R^2$  indicators are typically much smaller compared to these same statistics for aggregate level data from units such as counties or states.*

The multiple correlation here of  $r = .29$  indicates that the independent variables combined are weakly related to sentence length, and using the adjusted  $R^2$ , we can see that about 8% of the variation in sentence length is being explained by all the IV's combined. According to the model fit test of F, however, we see that this is model is explaining a significant amount of variation in sentences received by this sample of convicted homicide defendants.

Looking at the significance for each individual slope coefficient, we see that that there are only three variables that significantly affect sentence length received:

- Compared to female defendants, when defendants were male, they received an average of 3,291.5 days longer sentence days, even after controlling for the other IVs.
- Compared to defendants who had killed strangers, when defendants had killed someone they knew, the average sentence length was 1,977.1 days longer, even after controlling for the other IVs.
- Compared to those defendants who plead guilty, those who went to trial and were convicted received a sentence length that was about 8,873.2 days longer, which when divided by the 365 days of the year, is about 24.3 years longer, even after controlling for the other IVs.

Examining the t values here, we see that the type of adjudication, jury trial or plea, has the highest t value; in fact, it is nearly 4 times larger than the next highest t value. We can therefore conclude that the most important predictor of sentence length for convicted homicide defendants is whether the defendant plead or went to trial.

Let's predict the sentence length for a **29-year-old male white defendant who had no prior convictions, who killed one (single victim) individual he knew, and plead rather than go to a jury trial.** The appropriate values of  $x$  needed to calculate the predicted sentence length would look like this:

$$Y_{\text{sentence}} = 4999.12 + -5.9(29) + 97.1 (1) + 3291.5(1) + 1977.1(1) + 1440.8(0) + 8873.2(0)$$

$$Y_{\text{sentence}} = 4999.12 + -171.1 + 97.1 + 3291.5 + 1977.1 + 0 + 0$$

$$Y_{\text{sentence}} = \mathbf{10,193.7 \text{ days or } 27.9 \text{ years}}$$

How does this compare to the predicted sentence length for a similar **29-year-old male white defendant who had no prior convictions, who killed one (single victim) individual he knew, but who was convicted by a jury trial.** The only value of  $x$  we would change for this prediction would be for the variable "jurytrial" and the appropriate values of  $x$  needed to calculate the predicted sentence length would look like this:

$$Y_{\text{sentence}} = 4999.12 + -5.9(29) + 97.1 (1) + 3291.5(1) + 1977.1(1) + 1440.8(0) + 8873.2(1)$$

$$Y_{\text{sentence}} = 4999.12 + -171.1 + 97.1 + 3291.5 + 1977.1 + 0 + 8873.2$$

$$Y_{\text{sentence}} = \mathbf{19,066.9 \text{ days or } 52.2 \text{ years}}$$

This illustrates very clearly the effect of type of adjudication on sentence length. We see that a similar defendant with a similar homicide context and criminal background receives a sentence of over 24 years longer sentence compared to those who pled guilty.