This homework consists of two multiple-part questions, and is due Friday, February 21st, by 5pm (you can turn it in either in class, to my mailbox, to the TAs mailboxes, or in my office). Late homeworks will be penalized one letter grade per day (with homeworks received after 5pm that day counted as late). Due to the large enrollment in our class (approximately 90 people), we will not accept homeworks by email. Homeworks should be roughly 2 typed, double-spaced pages in length.

To complete this homework you’ll need to use SPSS. You will probably need to do this in the computer lab (the same place where we hold section each week). The data for this homework is on our class website. It consists of 9 variables observed across all 50 US states in 2010. The variables are (1) the name of the state, (2) the number of murders per 100,000 people, (3) the number of people incarcerated per 100,000 people, (4) the state population, (5) the percentage of people below the poverty line, (6) the percentage of people unemployed, (7) the number of foreclosures per 1000 homes, (8) the average summer temperature, (9) the percentage of people with a bachelor’s degree, and (10) the corporate tax rate.

**Question 1:** Suppose you were interested in testing the hypothesis “The higher the percentage of people in poverty in a state, the higher the murder rate in that state.”

- What variable in the data should you use for your dependent variable? What variable should you use for your independent variable?
- Using a histogram, examine your dependent and independent variables. Which observation or observations are outliers in your dependent variable (if any)? Which observation or observations are outliers in your independent variable (if any)?
- Calculate the correlation between your dependent and independent variables. What is this correlation? Based on this correlation, does it appear there is support for the hypothesis?
- Run a regression to examine the relationship between the dependent and independent variables. What is the intercept of the regression line, and what does it tell you in this case? What is the slope of the regression line, and what does it tell you in this case? Based on this regression line, does there appear to be support for the hypothesis?
- Generate a scatterplot with a regression line showing the relationship between your dependent and independent variables (make sure your dependent variable is on the vertical axis). Print out this scatterplot and attach it to your homework.
**Question 2:** Now generate your own original hypothesis about a relationship between two variables in these data. Make sure this isn’t a hypothesis we’ve already examined in section or in Question 1.

- State your hypothesis. In a sentence or two, explain why you expect to see this relationship.

- What is the dependent variable in your hypothesis? What is the independent variable?

- Using a histogram, examine your dependent and independent variables. Which observation or observations are outliers in your dependent variable (if any)? Which observation or observations are outliers in your independent variable (if any)?

- Calculate the correlation between your dependent and independent variables. What is this correlation? Based on this correlation, does it appear there is support for your hypothesis?

- Run a regression to examine the relationship between the dependent and independent variables. What is the intercept of the regression line, and what does it tell you in this case? What is the slope of the regression line, and what does it tell you in this case? Based on this regression line, does there appear to be support for your hypothesis?

- Generate a scatterplot with a regression line showing the relationship between your dependent and independent variables.

- Print out your SPSS output for this question (the histograms, the correlation, the regression, and the scatterplot) and attach it to your homework.