

Stata Assignment #3

1. Consider the relationship between women's education (EDUC) and income using the 2018 GSS data. Select women for whom information is available regarding income and years of schooling completed (code as missing those without income as well as those for whom there is no information). Recode income (RINCOM16) using the midpoints for all but the top category; assign (arbitrarily) \$190,000 for the top category. For example, for category 2, assign the value of \$2,000; for category 9, assign the value \$11,250. Regress income on years of school completed. Make a table reporting your regression results. Interpret the coefficient. Is it statistically significant? How much of the variance in income is explained by education?
2. Make a new variable containing the predicted values by using the command `-predict-`. What is the expected income of a female high school graduate? What is the expected income of a female with a B.A. (i.e., four years past high school)? Or of a female with two years of graduate or professional school?
3. What is a possible confounding variable in the relationship between education and income you observe? Test at least one possible confounding variable that can be tested with the GSS using multiple regression. Make a table reporting your regression results with the two nested models. Interpret the coefficients. Are they statistically significant? Do you continue to observe a relationship between education and income among women? Do you explain more of the variance in income with this additional variable? [Note: Make sure that you have the same sample size for these two regressions; if you do not, adjust your sample size in the first regression to account for the missing cases of the added variable(s)].
4. Choose any two variables from the GSS. Propose a theory relating the variables that can be tested by estimating a regression equation. Recode or transform the variables as necessary, estimate the equation. Describe your results by interpreting your coefficients and indicating what they imply about your theory. Now consider a possible confounding variable or variables and include the additional variables in your analysis. Present a table of regression results of the two nested models. Describe the results, including the change in the key independent variable that results from the inclusion of the additional independent variable(s).
5. Include your log file of all the calculations for this assignment.