

PUBPOL 639: ASSIGNMENT 2

Winter 2011

Due: Monday, February 21st at the start of class

This assignment aims to get you comfortable with running and interpreting the output from bivariate regression. Be sure to copy your do and log files into the back of your solutions. Courier 8 or 9 pt font works well. Your log file will be long, so don't worry about cleaning/formatting it.

PROBLEM 1 – School Construction as Economic Development

Governor Snyder has decided to make extending access to higher education a central component of his plan to revitalize Michigan's economy. He's considering either providing subsidies for students from low-income families or allowing some rural community colleges to begin offering bachelor's degrees. Underlying this latter initiative is the belief that geographic accessibility is an important determinant of college attendance. He asks you to assess the importance of family income and distance to the nearest Bachelors-granting institution to educational attainment. In this exercise you will investigate the relationship between the completed schooling of adults, the distance of their childhood homes to the nearest college, and family income. You will examine this relationship using data on a random sample of 1982 high school seniors who were re-interviewed six years after high school to determine how many years of education they had completed. The data and its documentation can be found on the CTools site (Resources > Data).

1. Describe the data

- a. Summary Statistics
 - i. How does completed schooling vary by family income?
 - ii. How does completed schooling vary by whether a student's mother graduated college? What about their father?
 - iii. Describe the variation in distance to college. What is the mean, median, min, max, 25th percentile, 75th percentile?
 - iv. The distance data are top-coded. What is the top code value? What is the largest fraction of the observations that could be top-coded?
- b. Graphical analysis
 - i. Graph completed schooling against distance to nearest college using a scatterplot. Include the graph in your answers.
 - ii. What do you see? Does there appear to be a relationship between distance to college and years of completed education?

2. Regression analysis I

- a. Regress completed schooling on distance to the nearest college
- b. Write out the equation that corresponds to this regression.
- c. Find and label the following on your Stata output and interpret these statistics.
 - i. $\hat{\beta}_1$
 - ii. Standard error of $\hat{\beta}_1$
 - iii. t-test for null hypotheses that $\hat{\beta}_1 = 0$

- iv. confidence interval for $\hat{\beta}_1$
 - v. $\hat{\beta}_0$
 - vi. Number of observations
 - vii. R^2
- d. Based on this regression, what is the predicted schooling of a person who lives ten miles from the nearest college? Fifty miles? Show your work.

3. Regression analysis II

- a. Regress completed schooling on the variable indicating whether the child comes from a high income family.
- b. Write out the equation that corresponds to this regression.
- c. What is the “omitted” category in this regression?
- d. Find and label the following on your Stata output and interpret these statistics.
 - i. $\hat{\beta}_1$
 - ii. Standard error of $\hat{\beta}_1$
 - iii. t-test for null hypotheses that $\hat{\beta}_1 = 0$
 - iv. confidence interval for $\hat{\beta}_1$
 - v. $\hat{\beta}_0$
 - vi. Number of observations
 - vii. R^2
- e. Based on this regression, what is the predicted schooling of a person who comes from a high income family? What about one that comes from a low income family? Show your work.

4. Model fit

Which factor – distance to college or family income – explains more of the variation in educational attainment in this data?

PROBLEM 2 – Bed Nets

The purpose of this exercise is to help you learn the mechanics of ordinary least squares (OLS) regression. First you will calculate the regression “by hand” then you will use Stata to confirm the calculation.

The table below contains data on the fraction of children enrolled in primary school and the fraction of children under five sleeping under insecticide-treated bed nets for five countries in 2007. Bed nets have been shown to be the most cost-effective prevention method against malaria and are part of UN’s Millennium Development Goals (which is where this data comes from).

	Primary school enrollment rate	Fraction of children sleeping under insecticide-treated bed nets
Ethiopia	72.3	33.1
Indonesia	98	3.3
Namibia	88.1	10.5
Swaziland	87.2	0.6
Zambia	95.4	28.6

- Using the appropriate formulas, show how to calculate each of the following. Note: “show how to calculate” means (1) write the appropriate formula; (2) plug in the appropriate values; and (3) show the computed answer. You do not need to show the intermediate calculations between steps 2 and 3. You may use Excel to do the calculations as long as you show the formulas used and you do not use the built-in regression function.
 - $\hat{\beta}_1$ - the estimated slope coefficient from the regression of enrollment rate on fraction sleeping under bed nets
 - $\hat{\beta}_0$ - the estimated intercept from the same regression
 - \hat{Y}_i - the predicted values for the five countries
 - \hat{u}_i - the OLS residuals for the five countries
- In a concise paragraph drawing on the numbers you calculated above, describe the relationship between bed net utilization and the primary school enrollment rate as precisely as you can. Indicate the direction and magnitude of the relationship based on this sample of five countries.
- Now you will see how the same regression is produced by Stata. Open Stata and type “edit,” which brings up something that looks like a spreadsheet. Enter the country name, enrollment, bed net entries in the first three columns. Double-click the column headers to enter variable names (“country”, “enrollrate”, “bednet”). Close the editor window when you are done. Type “list” to be sure you have typed in the numbers correctly, and type “sum” to inspect the variable means. Regress the enrollment rate against the bed net utilization rate. On the Stata output, find and label $\hat{\beta}_0$ and $\hat{\beta}_1$.

4. Depending on the sign of the association you find, answer the appropriate question below (i.e. only answer a or b):
- a. If you found a positive association, should we interpret your finding as evidence that bed nets improved primary school enrollment? Why or why not.
 - b. If instead you found a negative association, should we interpret your findings as evidence that bed nets decrease primary school enrollment (or at least don't help)? Why or why not.