Homework 2 - ECON 5453

Travis Roach

September 13, 2019

1.

The Department of Economics at Metrics University randomly sampled 22 MSBA students at the beginning of the Fall 2018 semester and obtained data on the following variables: the current grade point average of student "i" (call this variable "GPA_i"); the grade point average of the student upon graduating from their undergraduate (call this variable "$UGPA_i$"); the average number of hours per week that student i spent at the bar (call this variable "BAR_i"); and the average number of hours per week that student i studied (call this variable "$STUDY_i$"). The Department of Economics at Metrics University estimated (by OLS) three alternative regression models using these variables, and their results are shown below:

(1.) $GPA_i = 0.90 + 0.5UGPA_i + \hat{u}_i \quad r^2 = 0.2381$

(2.) $GPA_i = 2.0 - 0.1BAR_i + \hat{u}_i \quad r^2 = 0.1667$

(3.) $GPA_i = 1.6 + 0.3STUDY_i + \hat{u}_i \quad r^2 = 0.3103$

(a) Explain the economic interpretation of the estimated slope coefficients for each of the three models.

The marginal effect of a unit change in undergraduate GPA is a 0.5 unit change in current GPA; The marginal effect of a unit change in the amount of hours spent in the bar each week is a -0.1 unit change in current GPA; The marginal effect of a unit change the amount of hours studied in a week is a 0.3 unit change in current GPA

(b) Explain the interpretation of the numerical values of the r^2 statistics for each of the three models. Then, for each model, explain whether there is a relatively strong or relatively weak relationship between X and Y (support your answers).

The strongest relationship is found in regression 3 while the weakest relationship is regression 2. For cross-sectional data, an r^2 that is below 0.20 generally indicates poor model fit. *The r^2 statistic indicates the amount of variation in GPA that is attributable to variation in the independent variable(s). For example, 23.81% of the variation in GPA can be explained by the variation in undergraduate GPA.*

The strongest relationship is found in regression 3 while the weakest relationship is regression 2. For cross-sectional data, an r^2 that is below 0.20 generally indicates poor model fit.

2.

You have been hired by the G Markov Consulting Group to analyze the relationship between automobiles and air pollution around the globe. In particular, they want to know how the number of cars in the world affects the amount of air pollution, and how strong this relationship is. To that end, you have collected data and run a simple regression model of the form:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

Where Y_i represents the number of units of pollution in country i (per cubic foot of air), and X_i represents the number of cars in country i (measured in hundreds of thousands of cars).

The following coefficients have been estimated for this equation:

$$\hat{\beta}_0 = 2.92,\ \hat{\beta}_1 = 0.52;\ r^2 = 0.5217$$

(a) How many units of pollution will there be if there are no cars in a given country?

2.92

(b) How will the amount of pollution change if the number of cars increases by one unit?

0.52

(c) In Oklahoma alone there are approximately 1.37 Million registered vehicles. If all of these vehicles were instead electric vehicles, how much would emissions change by?

This is tantamount to removing 1.37 Million vehicles from the roads in the United States. If that were to occur, our model predicts that there will be 7.124 fewer units of pollution (-13.7 *unit change in cars* $*0.52 = -7.124$*)*

(d) The G Markov consulting group would like to know if there is a relatively strong or weak relationship between units of pollution and number of cars. Discuss using relevant statistic(s).

In this model, approximately 51.17% of the variation in pollution is explained by variation in the amount of cars on the road. This value for r^2 *indicates a relatively strong relationship.*

3.

State and explain the assumptions that are necessary for the ordinary least squares estimates to be:

(a) unbiased and consistent.

(b) efficient.

(c) unique (that is, in order for these estimates to exist).

See class notes for solutions to a-c.