

ENV ECON 118 / IAS 118 - Introductory Applied Econometrics
Assignment 4
Due Tuesday November 3 at the beginning of class

1. Submit a clean stata log file
2. Please report at least 3 decimal places
3. Please write your answers clearly and concisely

Exercise 1: Fuel Consumption

Background

The data file *PS4_cars* contains observations for 300 cars on their miles per gallon (*mpg*), their number of cylinders (*cyl*), their engine displacement (some technical parameter of the car related to the size of the piston and the number of cylinder), horsepower (*hp*), and the vehicle weight in pounds (*wght*).

1. Estimate the model:

$$mpg = \beta_0 + \beta_1 hp + \beta_2 wght + u$$

- (a) Report the results (write out the estimated equations)
 - (b) Comment on your results (SSS for both slope parameters)
 - (c) Compare the relative effect of weight and horsepower in determining fuel efficiency (hint: think standardizing).
2. Use your results above to:
 - (a) Predict the average fuel efficiency (mpg) of the Mazda Tribute SUVs i.e. a car that weighs 3500lbs with an engine of 240 hp.
 - (b) Run a regression that allows you to put a 95% confidence interval around this predicted value.
 - (c) When you interpret your results, do so in a way that is both precise and meaningful.
 3. Let mpg^0 be the unknown fuel efficiency of one particular SUV that weighs 3500lbs with an engine of 240 hp.
 - (a) Find a 95% confidence interval for mpg^0 .
 - (b) Compare the width of this confidence interval with what you found in the previous question. Explain.
 4. Plot the graphs of fuel efficiency (*mpg*) against (*hp*) and against weight. What functional form do they suggest might be more appropriate than a linear form for explaining fuel efficiency?
 5. Now estimate the following model

$$\ln(mpg) = \beta_0 + \beta_1 hp + \beta_2 wght + u$$

- (a) Report the results (write out the estimated equation)
- (b) Find the predicted value of *mpg* for that same Mazda Tribute SUV.

6. For explaining variation in *mpg*, decide whether you prefer the model from question 1 or the model of question 5.
7. Now use these results to make a policy statement. Whichever position you take, you need to support your argument with convincing quantitative results. Be creative. Lets discover your talents for advocacy!

Exercise 2: Do legislators with more daughters vote more liberally on women's issues?

Background

The data for this exercise were used in Ebonya Washington's paper: "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues." published in the American Economic Review in 2008. The paper asks whether having daughters influences the voting behavior of members of the US Congress. The hypothesis is that having (more) daughters makes legislators more likely to vote liberally on issues concerning women.

For this exercise, we will focus on votes that took place in the 108th Congress, which held session in 2003/04. As a measure of a liberal voting record, we use scores assigned by the American Association of University Women, a liberal group that concerns itself with issues of interest to women. For the 108th Congress, the AAUW selected 9 pieces of legislation in the areas of education, equality and reproductive rights. The AAUW then assigned a score to each member of Congress. The scores range from 0 to 100 and measure the percentage of times the legislator voted in favor of the position held by the AAUW.

The dataset *PS4.Legislators* contains the following characteristics of members of the 108th Congress:

- *ngirls* number of daughters
- *totchi* number of children
- *age* Age
- *female* indicator for being female
- *repub* indicator for being a Republican
- *moredef* proportion of people in the legislator's district who are in favor of "more spending on defense"
aauw AAUW score

A few useful stata commands (You can use the help stata command for more details)

- **predict newvar,xb** If used after a regression, will create a variable *newvar* with the predicted value $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$
- **correlate var1 var2** Prints the matrix of correlation between the two variables var1 and var2
- **test $x_1 + x_2 = 1$** Tests linear hypotheses after estimation, in this case the hypothesis $H_0 : \beta_1 + \beta_2 = 1$ on the parameters corresponding to x_1 and x_2 (note that you use the names of the variables to indicate the corresponding parameters)
- **gen ngirl1=ngirls==1** Creates a dummy variable equal to 1 if *ngirls* = 1, and 0 otherwise

1. First examine the data and give a brief description of the legislators in the data base. In particular:
 - (a) Look at the proportion of male and female legislators
 - (b) Look at the distribution of *ngirls* (use a histogram)
 - (c) What is the average age of legislators

2. Estimate the following regression models:

$$aaaw = \beta_0 + \beta_1 ngirls + u \quad (1)$$

$$aaaw = \beta_0 + \beta_1 ngirls + \beta_2 totchi + u \quad (2)$$

$$aaaw = \beta_0 + \beta_1 ngirls + \beta_2 totchi + \beta_3 female + \beta_4 repub + \beta_5 moredef + u \quad (3)$$

Please report your results (write out the estimate equations)

3. Interpret the result of the above regressions. In particular:

- (a) Briefly interpret all parameters of the 3rd model (remember SSS).
- (b) Discuss how the coefficient on β_1 changes across the three models. What do you think explains the difference in β_1 between the first and second model?

4. Is liberal voting correlated with age? Estimate a linear and a quadratic regression of voting scores on age. Analyze your results.

5. Start from model (c), in which we analyze whether the voting pattern of a legislator is influenced by the number of girls s/he has. Modify this model to explore three other ways to account for the influence of having girls on voting pattern:

- (i) having at least one girl (you need to create a dummy variable equal to 1 if the legislator has at least one girl and 0 if the legislator has no girl)
- (ii) having more than one girl (you need to create two dummy variables, one for having exactly one girl and one dummy for having more than one girl)
- (iii) the share of girls among the children

Now,

- (a) Report your results for each equation (write out the estimated equation)
- (b) Which of these models gives the best representation of the influence of girls on voting?
- (c) Interpret the results of the best model (discussing SSS for the girls parameters only).

6. Start from model (c) again. Now we would like to know whether the effect of having daughters is different for male and female legislators. Estimate a model that will allow you to test this, and interpret your results.

7. Estimate model (c) separately for male and female legislators. Formally test if the regression should be run separately or if it's ok to pool the data.