Name:_____

ECON 453
In-Class Exercise 5
October 5, 2023

Please download the file "IC5 Session.gretl", a gretl Session file. This dataset comes from the 2019 American Community Survey and is very similar to what we are using in Problem Set 2. The data contain information at the individual level on work outcomes and demographic information. The version of the dataset we are looking at today includes individuals that have an Economics undergraduate degree (some have master's as well), work at least 30 hours per week, make at least $15,000 per year, and are between the ages of 25 and 45. Please open the session file. The dataset contains basic descriptions of each of the variables.

1. Let's have some fun with OVB.
    a. Regress income on the following regressors: female and hours worked. Report and briefly interpret the female coefficient.

    b. How biased might our estimate of the gender gap be if we do not know the hours of work for an individual (we were forced to omit it from our model). To find this, you first need to run a regression where hours is the dependent variable and female is the independent variable. Report this coefficient and discuss what it tells us.

    c. Use the estimated coefficients on hours from part a and female from part b to determine how much bias we would have if we omitted hours of work from our model estimating the gender gap. What do you expect the estimated gender gap would be after the bias?

    d. Run a regression where you regress income on (only) the female dummy variable. Compare to your estimate from part c and be impressed (or sad).

2. Run a regression where you use income as the dependent variable and the following regressors: female, age, hours, masters.
    a. Briefly discuss the findings and whether these match your expectations.

    b. Find/calculate the standardized coefficients for the hours and age variables from the regression in part a. Comment briefly on what these tell us.

c. Run the same model (as in part b) but add dummy variables for race categories. Think carefully about how you create categories/adjust the sample. You should include at least two dummy variables for race/ethnicity. Briefly explain your thought process on how to add race to the model. Briefly summarize what you learned about race and income.

d. Use the results in your model from part c. Test the equality of two (or more!) of the coefficients Report what you tested and your conclusion (briefly).

3. For this set of models, we will use hours worked per week (**hours**) as the dependent variable.
   a. Create two dummy variables, one for "married", and one for "kids". The "married" should be a simple yes/no regarding whether the person is currently married. The "kids" variable should be 0 if the person has no children in the household and 1 if there are any. Explain briefly how you created these.

   b. Run a regression using hours as the dependent variable where regressors are female, age, married, kids. Briefly comment on your findings.

   c. Create an interaction term between married and female. Add this to the model from part b. Discuss what this tells us about gender, marriage, and hours worked.

   d. Now create a second interaction term between kids and female. Add this to the model from part c (yes, you will have two interaction terms!). Create two separate prediction equations, one for males and one for females. The equations should tell us how to predict hours worked based on age, marital status, and child ownership. Summarize what we learn.