Name:_____

ECON 453
In-Class Exercise 6
October 19, 2023

Please download the file "IC6.gdt", a gretl data file. This dataset comes from the 2019 American Community Survey and is similar to the dataset we used in Problem Set 2. The data contain information at the individual level on work outcomes and demographic information. The dataset we are looking at today includes individuals that have a bachelor's degree in economics, accounting, marketing, or finance, work at least 30 hours per week, make at least $15,000 per year, and are between the ages of 25 and 40. Please open the data file. The dataset contains basic descriptions of each of the variables.

1. We are going to work with a binary dependent variable, whether or not the individual has ever been married in their (short) life. Please define this variable using the "marst" variable in the data.
    a. Run a linear probability model (OLS regression) using the "ever married" variable as the dependent and the following regressors: female, age, immig.
        i. Interpret the coefficients and discuss briefly whether these make sense.

    b. Run the model again and add the "nchild" variable as an additional regressor.
        i. Interpret the coefficient on the nchild variable
        ii. Predict the probability that Janky McMurphy is/was married. She is a 33-year-old female that immigrated from Ireland. Janky has 4 kids and a heart of gold. Comment on your prediction.
        iii. Was the addition of nchild an improvement to our model?

    c. Save the fitted values from your regression in part b (from the results menu, select **Save -> Fitted values**). Generate the summary statistics and frequency distribution for these fitted values.
        i. Comment on what these tell us about our estimated probabilities.

d. Now run the same regression using a binary Logit model (instead of OLS). Use the same regressors as in part b.
  i. Compare the predicted impacts of each variable to what we saw in part b.
  ii. Summarize what we learn from the "% correctly predicted" table in the results. What percentage of the sample has been married? What percentage of the sample does our model predict has been married?

e. Save the fitted values from your regression in part d. Generate the summary statistics/frequency distribution for these fitted values.
  i. Comment on how these compare to what we saw in part c.

f. Using your model in part d, what is your predicted probability that Janky McMurphy has been married