

Homework 11

MRM Qualitative Regressors, Part 2 (10 points)

Instruction:

- This HW must be done in Rmarkdown!
- Please submit both the .rmd and the Microsoft word files. (Do not submit a PDF or any other image files as the TAs are going to give you feedback in your word document)
- Name your files as: HW10_groupnumber_name
- All the HW assignments are individual work. However, I highly encourage you to discuss it with your group members.
- The answer key will be uploaded on Canvas a couple of days after the due date.

Question 1 Given the following:

$$\begin{aligned} inlf = & 0.586 - 0.0034nwifeinc + 0.038educ + 0.039exper - 0.0006exper^2 \\ & - 0.016age - 0.262kidslt6 + 0.013kidsage6 \end{aligned}$$

$$n = 753, R^2 = 0.264$$

The variables are defined as:

inlf = (binary variable) In labor force participation by married women.

nwifeinc = Husband's earnings measured in thousands of dollars.

educ = Years of education.

exper = Years of labor market experience.

age = Age.

kidslt6 = Number of children under age 6.

kidsge6 = Number of kids between 6 - 18 years old.

Suppose that we define *outlf* to be one if the woman is out of the labor force, and zero otherwise.

1. If we regress *outlf* on all of the independent variables in the equation, what will happen to the intercept and slope estimates? (Hint: $inlf = 1 - outlf$. Plug this into the population equation $inlf = \beta_0 + \beta_1nwifeinc + \beta_2educ + \dots$ and rearrange.)
2. What will happen to the standard errors on the intercept and slope estimate?
3. What will happen to the R-squared?

Computer Exercises

Question 2 Use the data in APPLE for this exercise.

(Hint: in order to make a new dummy variable called *ecobuy*, you need to use a combination of *mutate()* and *ifelse()* functions.)

1. Define a binary variable as *ecobuy* = 1 if *ecolbs* > 0 and *ecobuy* = 0 if *ecolbs* = 0. In other words, *ecobuy* indicates whether, at the prices given, a family would buy any ecologically friendly apples. What fraction of families claim they would buy ecolabeled apples?
2. Estimate the linear probability of the following model.

$$\begin{aligned} ecobuy = \beta_0 + \beta_1 ecoprc + \beta_2 regprc + \beta_3 faminc + \beta_4 hhsz \\ + \beta_5 educ + \beta_6 age + u \end{aligned}$$

Report the results in equation form. Carefully interpret the coefficients on the price variables.

3. Are the nonprice variables jointly significant in the LPM? (Use the usual F statistic, even though it is not valid when there is heteroskedasticity.) Which explanatory variable other than the price variables seems to have the most important effect on the decision to buy ecolabeled apples? Does this make sense to you?
4. In the model from part 2, replace *faminc* with *log(faminc)*. Which model fits the data better, using *faminc* or *log(faminc)*? Interpret the coefficient on *log(faminc)*.
5. In the estimation in part 4, how many estimated probabilities are negative? How many are bigger than one? Should you be concerned?
6. For the estimation in part (iv), compute the percent correctly predicted for each outcome, *ecobuy* = 0 and *ecobuy* = 1. Which outcome is best predicted by the model?