

# Extra Practice Questions - Week 3

## Statistics and Econometrics

### Question 1

Use the data in `apple.RData` to answer this question.

1. Define 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 fractions of families claim they would buy eco-labeled apples?
2. Estimate the linear probability model

$$ecobuy = \beta_0 + \beta_1 ecoprc + \beta_2 regprc + \beta_3 faminc + \beta_4 hhsz + \beta_5 educ + \beta_6 age + u,$$

and report the results. Carefully interpret the coefficients on the price variables.

3. Are the nonprice variables significant in the LPM? Which explanatory variable other than the price variables seems to have the most important effect on the decision to buy eco-labeled apples? Does this make sense to you?
4. In the estimation in part 2, how many estimated probabilities are negative? How many are bigger than one? Should you be concerned?

### Question 2

The variable *smokes* is a binary variable equal to one if a person smokes, and zero otherwise. We estimate a linear probability model for smokes:

$$\begin{aligned} \widehat{smokes} = & \begin{matrix} .656 & - & .069 & \log(cigpric) & + & .012 & \log(income) & - & .029 & educ \\ (.855) & & (.204) & & & (.026) & & & (.006) \\ [.856] & & [.207] & & & [.026] & & & [.006] \end{matrix} \\ & + \begin{matrix} .020 & age & - & .00026 & age^2 & - & .101 & restaurn & - & .026 & white, \\ (.006) & & & (.00006) & & & (.039) & & & (.052) \\ [.005] & & & [.00006] & & & [.038] & & & [.050] \end{matrix} \end{aligned}$$

$n = 807$ ,  $R^2 = .062$ . *income* is the person's annual income; *cigpric* indicates the per pack price of cigarettes; *educ* indicates years of schooling the person received; *age* is measured in years; *restaurn* is a dummy variable which equals one if the person lives in a state with restaurant smoking restrictions; *white* equals to one if the person is Caucasian. Both the usual and heteroskedasticity-robust standard errors are reported.

1. Are there any important differences between the two sets of standard errors?
2. Holding other factors fixed, if education increases by four years, what happens to the estimated probability of smoking?
3. Interpret the coefficient on the binary variable *restaurn*.
4. A person in the data set has the following characteristics:  $cigpric = 67.44$ ,  $income = 6,500$ ,  $educ = 16$ ,  $age = 77$ ,  $restaurn = 0$ ,  $white = 0$ , and  $smokes = 0$ . Compute the predicted probability of smoking for the observation in the data set.