Experiential Learning
. Binary-Choice Model. Spring 2018.
Due Date: 02-26-2019.

Rules:

1. You have to present your results in classroom by using your power-point slides.

2. Group work, with at most 3 students per group, is expected but not required.

3. You will have to turn in your power point slides, do file and answer sheet on 02/26/2019 at 10:55 AM.

4. I will grade not only your answers to questions below but also your answers to questions I will ask during your presentation.

5. Group work means that each individual in the group is responsible for the grade of the whole group. Wrong/Right answers during presentation in classroom will penalize/benefit all members of the group.

We want to analyze the purchase of private insurance (ins). We consider the following explanatory variables:

1. hstatusg: a dummy variable that measures whether health status is good or not.
2. married: marital status
3. hisp: dummy variable equal to 1 if hispanic
4. hhincome: household income
5. retire: dummy variable equal to 1 if retired
6. educyear: years of education.
7. age
8. sretire: dummy variable equal to 1 if retired spouse is present.
9. linc: log household income
10. female: dummy variable equal to 1 if female
11. white: dummy variable equal to 1 if white

You will use the data musdata.dta to answer the following questions.

Question 1. Use the summary command to summarize your dataset. Use the global command to create a variable list called "xlist" that includes age, hstatusg, hhincome, educyear, married and hisp.

Question 2. Use the logit command to estimate a logit model in which the dependent variable is the purchase of private insurance (ins) and the explanatory variables are the ones that you included in "xlist" . Interpret your results.

Question 3. Do same with the probit command. Interpret your results

Question 4. Write the following code:

.*estimation of several models

. quietly logit ins retire $xlist

. estimates store blogit
. quietly probit ins retire $xlist
. estimates store bprobit

This will lead to the following output table of parameter estimates across the two models

. * Table for comparing models
. estimates table blogit bprobit, t stats(N ll)

Interpret the results.

Question 5. Now let's use the Wald test. You must interpret your results.

.* Wald test for zero interactions
. generate age2=age*age
. generate agefem=age*female
. generate agechr=age*chronic
. generate agewhi=age*white
. global intlist age2 agefem agechr agewhi
. quietly logit ins retire $xlist $intlist
. test $intlist
(1) [ins]age2=0
(2) [ins]agefem=0
(3) [ins]agechr=0
(4) [ins]agewhi=0

Question 6. Now you will use the likelihood ratio test for testing the same restrictions (null hypothesis) as in Question 5. Interpret your result and compare it to the one obtained in question 5.

. * Likelihood-ratio test
. quietly logit ins retire $xlist $intlist
. estimate store B
. quietly logit ins retire $xlist
. lrtest B

Could you test the same null hypothesis by using the F-test rather than Wald and LR ? Explain your answer.

Question 7. Suppose you want to predict whether an individual has private insurance ($\widehat{y} = 1$) or does not ($\widehat{y} = 0$). Let's assume that $\widehat{y} = 1$ if $F\left(x'\widehat{\beta}\right) > 0.5$ and $\widehat{y} = 0$ if $F\left(x'\widehat{\beta}\right) \leq 0.5$. Our measure of goodness of fit is the percentage of correctly classified observations. Use the estat classification command to obtain the percentage of correctly classified observations. In other words, type the following code

.* Comparing fitted probability and dichotomous outcome
. quietly logit ins retire $xlist
. estat classification

How many observations are misclassified as 1 when the correct classification is zero? How many observations are misclassified as 0 when the correct value is 1. How many observations are correctly classified ?

Question 8. This question is about computation of marginal (partial) effects. We are going to consider three options: marginal effect at a representative value (MER), marginal effect at the mean (average) (MEM); and average marginal effect (AME)

(i) The postestimation *margins* command with the dydx() and at() options provides an estimate of the marginal effect at a particular value of $\mathbf{x}=\mathbf{x}^*$. In our empirical application we will use as a benchmark a 75-year-old, retired, married Hispanic with good health status, 12 years of education, and an income equal to 35. For the four binary regressors, you are asked to compute the MEM using the finite-difference method, rather than the calculus method, by using the **i.operator** in the logit estimation command. In other words,

.*Marginal effects (MER) after logit
. quietly logit ins i.retire age i.hstatusg hhincome educyear i.married i.hisp
. margins, dydx(*) at (retire=1 age=75 hstatusg=1 hhincome=35 educyear=12
> married=1 hisp=1) noatlegend // (MER)

You are exptected to interpret your results.

(ii) For comparison, you now compute the marginal effect at the mean, using the *margins* command with the *atmean* option. In other words,

. * Marginal effects (MEM) after logit
. quietly logit ins i.retire age i.hstatusg hhincome educyear i.married i.hisp
. margins, dydx(*) atmean noatlegend // (MEM)

Interpret your results. How different they are from the results obtained in (i) ?

(iii) The average marginal effect (AME) is obtained as the default option of the *margins, dydx*() command. The associated standard errors and confidence interval for the AME are obtained using the delta method. Type the following commands:

. * Marginal effects (AME) after logit
. quietly logit ins i.retire age i.hstatusg hhincome educyear i.married i.hisp
. margins, dydx(*) noatlegend // (AME)

Interpret your results.