

Binary Outcome Models

Instructions:

- Include .do file and log for this assignment
 - Turn in your answers as a word document or PDF (if you use LaTeX).
 - Include screen shots of regressions or better yet include tables (outreg2) with your assignment
-

1. Use ACP.dta
 - a. Create a binary variable (arrests2012) such that 1 = person arrested at least once in 2012, and 0 otherwise
 - b. Estimate a LPM relating arrests2012 to ppc, msent, tottime, ptime2012, inc2012, black, Hispanic, and born86
 - i. Report usual and robust standard errors
 - ii. What is the estimated effect on the probability of arrests if the variable ppc goes from 0.30 to 0.80.
 - iii. Test the joint significance of msent and tottime, using a nonrobust and robust test
 - c. Estimate a probit model relating arrests2012 to ppc, msent, tottime, ptime2012, inc2012, black, Hispanic, and born86.
 - i. At the average values of msent, tottime, inc2012, and ptime2012 in the sample, and with black = 0, Hispanic = 1, and born86=1, what is the estimated effect on the probability of arrests if the variable ppc goes from 0.30 to 0.80. Use the margins command in Stata
 - ii. Compare this result to the answer from part b(ii)
 - iii. Obtain the percent correctly predicted. What is the percent correctly predicted when arrests2012=0? When arrests2012=1?
 - iv. What do you make of these findings?
 - v. In the probit model, add the terms ppc-squared, ptime2012-squared and inc2012-squared. Are they jointly significant?
 - vi. Describe the estimated relationship between the probability of arrest and ppc (this includes the interpretation of ppc-squared).
 - vii. At what point does the probability of conviction have a negative effect on probability of arrest?
 - d. Estimate a logit model relating arrests2012 to ppc, msent, tottime, ptime2012, inc2012, black, Hispanic, and born86
 - i. At the average values of msent, tottime, inc2012, and ptime2012 in the sample, and with black = 0, Hispanic = 1, and born86=1, what is the estimated effect on the probability of arrests if the variable ppc goes from 0.30 to 0.80. Use the margins command in Stata.
 - ii. Compare this result to the answers from part b(ii) and c(ii).
2. Use affairs dataset from bcuse.
 - a. Describe your data and do the following:
 - i. Tabulate affair on naffairs. What do you notice?
 - ii. Tabulate affair on relig. What do you notice?
 - iii. Tabulate affair on ratemarr. What do you notice?
 - iv. Tabulate affair on yrsmarr. What do you notice?
 - b. Estimate a LPM relating affair to kids, male, yrsmarr, ratemarr using (i.), relig (using i.) with robust standard errors
 - i. Interpret the coefficients on kids, male, and yrsmarr. Explain simply how these affect the probability on affair.
 1. Are these individually significant?
 2. Are they jointly significant?

Binary Outcome Models

- ii. Interpret the results of the categorical `ratemarr` variable.
 1. Are they jointly significant?
- iii. What is the estimated effect on the probability of affair for a female with no kids with `ratemar=3`, who is anti-religion (`relig =1`), married for 5 years?
- iv. What is the estimated effect on the probability of affair for a female with no kids with `ratemar=3`, who is anti-religion (`relig =1`), married for 20 years?
- v. Compare (iii) and (iv)
- vi. Verify that the coefficient on `yrsmarr * 15 yrs` is equal to the difference between (iv) and (iii). Why is this the case?
- vii. Create a predicted affair variable and call it `yhat_ols`
 1. Summarize it. Explain the results
 2. Do they make sense? Explain
- c. Estimate a Probit model with the same dependent and independent variables.
 - i. What is the effect on the probability of affair for females with no kids, with `ratemarr=3`, who are anti religion, when the `yrsmarr` variable goes from 5 to 20? Use the margins command.
 1. Compare this result to 2(b)(v). Is it higher or lower? Why?
 - ii. Create a predicted affair variable and call it `yhat_probit`
 1. Summarize it
 2. Explain the results.
 - iii. Add a quadratic term for `yrsmarr` (`yrsmarr_sq`)
 1. Interpret the coefficient. What do you notice?
 2. Test the joint significance for `yrsmarr` and `yrsmarr_sq`.
 3. At what point do the years married have a negative impact on probability of affair?
 - a. Bonus – include a graph showing this
 - iv. Examine the margins for `relig` and `ratemarr`
 1. Use `marginsplot` to create a graph (include this as a .jpg in your assignment)
 2. Explain what the plot is saying. Does it make intuitive sense?
- d. Estimate a Logit model with same depending and independent variables as part b. (no need to include `yrsmarr_sq`)
 - i. Coefficients should be in log-odds term. Describe the coefficient of `yrsmarr` in relation to affairs
 1. Calculate the odds-ratio for `yrsmarr`
 2. What does this mean?
 3. Confirm that the odds-ratio you calculated is the same as the odds ratio you get for `yrsmarr` when you run “logit, or”
 - ii. What is the effect on the probability of affair for females with no kids, with `ratemarr=3`, who are anti religion, when the `yrsmarr` variable goes from 5 to 20? Use the margins command.
 1. Compare your results to the LMP and Probit models above
 - iii. Predict `yhat_logit`
 1. Summarize and explain
 2. How well does your model do in successfully predicting affairs?
 - a. What cut point do you use? Why? Explain your results.
- e. Write up a couple sentences on what you conclude after running LPM, Probit and Logit models regarding affairs and these variables. Was anything striking? Explain.