

Exercise - Week 6
A Binary Response Model

- Load the `wtp.dta` data set.
- Construct the log of income variable: **gen lny = ln(y)**
- The variable “EnvCon” measures the concern about the environment for the surveyed individual. It is equal to 1 if the individual is concerned and zero otherwise. Tabulate the values of the variable. In this exercise, we would like to relate the probability that one is concerned about the environment with a number of individual characteristics (age, sex, and income level) and one community characteristic (‘smell’ indicating whether there is a bad smell in the community).

Part 1: Linear Probability Model

- 1 Regress the concern on age, sex, log income and smell.
- 2 Which variables are statistically significant? What is the effect of the income on the probability of being concerned? Is it statistically different from zero at the 5% significance level?
- 3 Predict the residuals. Graph the distribution of the residuals. How does the distribution look?
- 4 Predict the probability of being concerned by the environment by typing **predict p0, xb**. Summarize the probabilities. In which range do they vary? Construct the mean of `p0` for each income groups by typing **egen mp0 = mean(p0), by(y)**.
- 5 Graph the environment concern and the predicted probabilities of a function of log income.

Part 2: Logit Models

- 1 Estimate the logit model by typing **logit EnvCon age sex lny smell**. How do you interpret the coefficient in front of log income? Is it statistically different from zero?
- 2 Predict the probabilities by typing **predict p1, p**. Summarize these probabilities. In which range do they vary? Construct the mean of `p1` for each income groups by typing **egen mp1 = mean(p1), by(y)**. Graph the environment concern and the predicted probabilities as a function of log income.
- 3 Estimate the logit model with the odds-ratio option by typing **logit EnvCon age sex lny smell, or**. Interpret the odds-ratio related to smell and to income.
- 4 Construct the odds-ratio for each individual ($or1_i = \Pr(Y_i = 1) / [1 - \Pr(Y_i = 1)]$). Compute the mean value for those who live in a bad neighborhood (smell = 1) versus a good one (smell = 0) by typing **su or1 if smell == 0**. How do you interpret these results in contrast with the previous question?

- 5 We want to test the null hypothesis saying there is no difference of environment concern according to age and sex. Specify the null and alternative hypotheses. Use the likelihood ratio test to make a conclusion with the 5% significance level (the critical value of chi-square distribution with 2 df is 5.99).

Part 3: Probit Models

- 1 Estimate the probit model by typing **probit EnvCon age sex lny smell**. How do you interpret the coefficient in front of log income?
- 2 Compare the estimated coefficients with the ones from the logit. Note that you can approximately deduce the logit ones by multiplying the ones estimated from the probit by 1.6.
- 3 Predict the probabilities by typing **predict p2, p**. Summarize these probabilities. In which range do they vary? Construct the mean of p2 for each income groups by typing **egen mp2 = mean(p2), by(y)**. Graph the environment concern and the predicted probabilities from each method (p0, p1, p2) as a function of log income.
- 4 We want to test the null hypothesis saying there is no difference of environment concern according to age and sex. Specify the null and alternative hypotheses. Use the likelihood ratio test to make a conclusion with the 5% significance level (the critical value of chi-square distribution with 2 df is 5.99).

Part 4: Marginal Effects of Income

- 1 Linear Probability Model: What is the marginal effect of income on the probability of being concerned? Construct a variable which contains the marginal effect calculated with the LPM for each individual (**gen me0 = 0.11208**).
- 2 Logit Model: Compute the marginal effect ($\partial \Pr(Y_i = 1) / \partial X_j = \beta_j \Pr(Y_i = 1) (1 - \Pr(Y_i = 1))$). What is the mean value? Graph the marginal effect as a function of log income.
- 3 Probit Model: Compute the marginal effect ($\partial \Pr(Y_i = 1) / \partial X_j = \beta_j \phi(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})$). To do this, first predict the value $\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$ by typing **predict xb2, xb**. Then construct the marginal effect of log income as **gen me2 = $\hat{\beta}_3$ *normd(xb2)**. Alternative, type **dprobit Evncon age sex lny smell**. What do you get? What is the mean value? Graph the marginal effect as a function of log income.
- 4 Graph all the marginal effects as a function of log income and compare them.