ECON 471

Fall 2020

Problem Set 5

Due Wednesday Dec 2, by Midnight CST

Name: Wenxiao Yang

Section: B3.

1. Use the data in LOANAPP.TXT for this exercise. The binary variable to be explained is *approve*, which is equal to one if mortgage loan to an individual was approved. The key explanatory variable is *white*, a dummy variable equal to one if the applicant was white. The other applicants in the data set are black and Hispanic.
To test for discrimination in the mortgage market, a linear probability model can be used:

$$approve = \beta_0 + \beta_1 white + other\ factors.$$

(i) If there is discrimination against minorities, and the appropriate factors have been controlled for, what is the sign of β_1?

positive.

(ii) Regress *approve* on *white* and report the results in the usual form. Interpret the coefficient on *white*. Is it statistically significant? Is it practically large?

$$\widehat{approve} = 0.70779 + 0.20060\, white$$

$$n = 1989 \qquad R^2 = 0.3201.$$

(iii) As controls, add the variables *hrat*, *obrat, , loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2*, and *vr*. What happens to the coefficient on *white*? Is there still evidence of discrimination against nonwhites?

decrease : new $\hat{\beta}_1 = 0.128820$.

$t = \frac{0.128820}{0.019732} > 1$ still significant

$\Rightarrow$ Yes. still the evidence.

(iv) Now, allow the effect of race to interact with the variable measuring other obligations as a percentage of income (*obrat*). Is the interaction term significant?

$t = \frac{-0.0042878}{0.0009453} > 1.$

significant.

(v) Now, estimate a probit model of *approve* on *white*. Find the estimated probability of loan approval for both whites and nonwhites. How do these compare with the linear probability estimates in part (ii)?

$\widehat{approve} = 0.54695 + 0.78395\, white$

whites : 100%

nonwhites : 54.695%.

(vi) Now, add the variables *hrat*, *obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2*, and *vr* to the probit model. Is there statistically significant evidence of discrimination against nonwhites?

Yes.

$t = \dfrac{0.520254}{0.096866} \doteq 5.371 > 1. \Rightarrow$ significant

(vii) Estimate the model from part (ii) by logit. Compare the coefficient on *white* to the probit estimate.

$\widehat{approve} = 0.8847 + 1.4094\, white.$

larger than probit estimate.

2. An equation explaining chief executive officer salary is

$$\widehat{\log(salary)} = 4.59 + .257\log(sales) + .011roe + .158finance + .181consprod$$

$$(.30) \quad (.032) \qquad\qquad (.004) \quad (.089) \qquad\qquad (.085)$$

$$-.283utility$$

$$(.099)$$

$$n = 209, \qquad R^2 = .357.$$

The data used are in CEOSAL1, where *finance*, *consprod*, and *utility* are binary variables indicating the financial, consumer products, and utilities industries. The base industry is transportation.

(i) Compute the approximate percentage in estimated salary between the utility and transportation industries, holding *sales* and *roe* fixed. Is the difference statistically significant at the 1% level?

-28.3%

$t = \dfrac{-0.283}{0.099} = -2.852$

significant

(ii) Use equation (7.10) in the textbook (ch 7) to obtain the exact percentage difference in the estimated salary between the utility and transportation industries and compare this with the answer obtained in part (i).

$100\% \times (e^{-0.283} - 1) = -24.64802\%.$

$> -28.3\%$

(iii) What is the approximate percentage difference in estimated salary between the consumer products and finance industries? Write an equation that would allow you to test whether the difference is statistically significant.

$$lsalary = \beta_0 + \beta_1 lsales + \beta_2 roe + \beta_3 consprod + \beta_4 utility + \beta_5 transportation + u$$

3. Let *grad* be a dummy variable for whether a student-athlete at a large university graduates in five years. Let *hsGPA* and *SAT* be high school grade point average and SAT score, respectively. Let *study* be the number of hours spent per week in an organized study hall. Suppose that, using data on 420 student-athletes, the following logit model is obtained:

$$\hat{P}(grad = 1|hsGPA, SAT, study) = \Lambda(-1.17 + .24hsGPA + .00058\,SAT + .073study),$$

where $\Lambda(z) = \exp(z)/[1 + \exp(z)]$ is the logit function. Holding *hsGPA* fixed at 3.0 and *SAT* fixed at 1,200, compute the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week.

$$-1.17 + 3 \times 0.24 + 1200 \times 0.00058 + 0.073\, study$$

$$\text{difference} = \Lambda(\cancel{7.346}\ 0.976) - \Lambda(\cancel{3.896}\ 0.611)$$

$$= \cancel{0.0193903}$$

4. Suppose you collect data from a survey on wages, education, experience and gender. In addition, you ask for information about marijuana usage. The original question is: “on how many separate occasions last month did you smoke marijuana?”

(i) Write an equation that would allow you to estimate the effects of marijuana usage on wage, while controlling for other factors. You should be able to make statements such as, “smoking marijuana five more times per month is estimated to change wages by *x*%.”

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 gender + \beta_4 marij + u.$$

(ii) Write a model that would allow you to test whether drug usage has different effects on wages for men and women. How would you test that there are no differences in the effects of drug usage for men and women?

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 gender + \beta_4 marij + \beta_5 drug + \beta_6 drug \cdot gender + u.$$

if $\beta_6 \neq 0$, there are differences.

(iii) Suppose you think it is better to measure marijuana usage by putting people into one of four categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times per month), and heavy user (more than 10 times per month). Now, write a model that would allow you to estimate the effects of marijuana usage on wage.

take the base group to be nonuser.

$$\log(wage) = \cancel{\beta_{g,0} + \beta_{g,1} educ + \beta_{g,2} exper + \beta_{g,3} gender + \beta_{g,4} marij_g}\ \beta_0 + \delta_1 lghtuser + \delta_2 moduser + \delta_3 hvyuser + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u.$$

~~$g = 1, 2, 3, 4.$~~

~~we estimate four categories sepately.~~

$H_0: \delta_1 = 0 \quad \delta_2 = 0 \quad \delta_3 = 0. \quad q = 3.$

$F_{3, n-8}.$

(iv) Using the model in part (iii), explain in detail how to test the null hypothesis that marijuana usage has no effect on wage. Be very specific and include a careful listing of degrees of freedom.

① test H_0:

$$\beta_{1,0} = \beta_{2,0} = \beta_{3,0} = \beta_{4,0} = \beta_0$$

$$\beta_{1,1} = \beta_{2,1} = \beta_{3,1} = \beta_{4,1} = \beta_1$$

$$\beta_{1,2} = \beta_{2,2} = \beta_{3,2} = \beta_{4,2} = \beta_2$$

$$\beta_{1,3} = \beta_{2,3} = \beta_{3,3} = \beta_{4,3} = \beta_3$$

$$\beta_{1,4} = \beta_{2,4} = \beta_{3,4} = \beta_{4,4} = \beta_4$$

$$F = \frac{\left(SSR_p - \sum_{g=1}^{4} SSR_g\right)/5}{\sum_{g=1}^{4} SSR_g / (n-20)} \sim F_{5, n-20}.$$

SSR_p is the SSR in (i)

② test. $H_0: \beta_4 = 0$. same in (i)

$$t = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)} \sim t_{n-5}$$