Econometrics
18/04/2013
Giuseppe Ragusa
Luiss - Department of Economics and Finance
PhD in Management

Problem Set

Please answer the questions on this sheet (or photocopy) in pen.

Name : _____

ID: _____

(Please type clearly)

1. The cost of attending your college has once again gone up. Although you have been told that education is investment in human capital, which carries a return of roughly 10% a year, you (and your parents) are not pleased. One of the administrators at your university/college does not make the situation better by telling you that you pay more because the reputation of your institution is better than that of others. To investigate this hypothesis, you collect data randomly for 100 national universities and liberal arts colleges from the 2000-2001 U.S. News and World Report annual rankings. You perform the following regression

$$\widehat{Cost} = \underset{(7311.17)}{(2058.63)} \times + \underset{(664.58)}{3985.20} \times Reputation - \underset{(0.13)}{0.20} \times Size + \underset{(2154.85)}{8406.79} \times Dpriv$$
$$- \underset{(1121.92)}{416.38} \times Dlibart - \underset{(1007.86)}{2376.51} \times Dreligion, \quad R^2 = 0.26$$

where Cost is Tuition, Fees, Room and Board in dollars; Reputation is the index used in U.S. News and World Report (based on a survey of university presidents and chief academic officers), which ranges students, and Dpriv, Dlibart, and Dreligion are binary variables indicating whether the institution is private, a liberal arts college, and has a religious affiliation.

The numbers in parentheses are heteroskedasticity robust standard errors.

(a) Are smaller colleges significantly more expensive than smaller college, everything else being equal?

(b) Test (at the 5% significance level) the null hypothesis that religiously affiliated college are not more or less expensive than non religiously affiliated college against a one sided alternative ($<$).

You also estimate the following model:

$$Cost = \underset{(1772.35)}{5450.35} + \underset{(590.49)}{3538.84} \times Reputation + \underset{(875.51)}{10935.70} \times Dpriv - \underset{(1180.57)}{2783.31} \times Dreligion, \quad R^2 = 0.23$$

(c) Use the information provided here along with any assumption you need to make to test whether the size of the college and the fact that it is a liberal art college are jointly statistically significant.

(d) Once `size` and `Dlibart` are removed from the initial regression, the estimated effect of attending a private institution has increased.

(2) The Survey on Household Income and Wealth (SHIW) began in the 1960s with the aim of gathering data on the incomes and savings of Italian households. Here we use data from the 2008 version which surveys about 8,000 households (24,000 individuals), distributed over about 300 Italian municipalities.

We use data on wages, education, and age of the individuals to study wage differentials among education attainments. The final sample we consider has data on 5889 working individuals between the age of 15 and 65.

We run the following regression:

$$\widehat{wage} = \underset{(1300)}{6856} + \underset{(348.3)}{2951} \times middle + \underset{(398.1)}{3932} \times middle.prof. + \underset{(373.4)}{6322} \times highschool + \underset{(741.5)}{6823} \times 3yrcollege$$
$$+ \underset{(613.4)}{10510} \times college + \underset{(1923)}{17690} \times graduate + \underset{(69.35)}{650.2} \times age - \underset{(.8795)}{4.970} \times age^2$$

where `wage` is the wage expressed in euro, `age` is the age of the individual expressed in years, and `elementary`, `middle`, `middle.prof`, `highschool`, `3yrcollege`, `college`, `graduate` are dummy variables denoting the highest level of education completed by the individual.

(a) What is the interpretation of the coefficient of the dummy `middle`?

(b) What is the wage premium of graduating from a 4yr college as opposed to stop after receiving a 3yr college degree? Is this effect statistically significant at the 5% significance level?

(c) Comparing two individuals with the same degree and both 32 years old, what will the increase in their salary be in three years?

(d) Look at the following STATA output:

```
. test 3yrcollege = college

 ( 1)  3yrcollege - college = 0

        F(  1,  5880) =    0.55
             Prob > F =    0.46
```

What is the null hypothesis being tested? (Describe the economic content of this hypothesis). Can you reject this null hypothesis at the 10% significance level?

3. For this exercise we will use the dataset Affairs.dta that contains infidelity data from a survey conducted by Psychology Today in 1969. The dataset contains observations for the following variables:

- *affairs*: How often engaged in extramarital sexual intercourse during the past year.
- *education*: numeric variable coding level of education:
  - 9 = grade school,
  - 12 = high school graduate,
  - 14 = some college,
  - 16 = college graduate,
  - 17 = some graduate work,
  - 18 = master's degree,
  - 20 = Ph.D., M.D., or other advanced degree;

- *gender*: dummy variable
  - =1 if female,
  - =2 if male;

You are asked to construct a statistical procedure to test whether male and female are not different when it comes to cheating conditional on the level of education.

(a) Run a linear probability model and carefully comment the output of the regression.

(b) What are the problem with a linear probability model?

(c) Re-run the analysis using either a logit or a probit model.

(d) Construct a statistical procedure to test whether females are more likely than males to cheat using the probit/logit and the linear probability model. Do the conclusions you reach using the two models agree? Are you surprised?

(e) Propose a procedure to test whether higher level of education (above the mean) is associated with higher propension to cheating.

4. For this exercise we use the dataset cig.dta that contains cross-section data on cigarette consumption on 46 U.S. States, for the year 1992. The dataset contains the following two variables:

price: avarage price of cigarette packs per state;

packs: cigarette consumption (in packs) per person of smoking age (¿16 years old).

(a) You are working at the Health Department. You want to know whether cigarette price has an effect on smoking habits, in particular, whether price hikes lead to a reduction in smoking. Do a regression and comment the output (Hint: *you have to comment carefully significance of coefficients, R^2, etc.*).

Based on the result of the regression, do you think you can reach a scientifically solid conclusion? Explain.

(b) If you express prices in Euro instead of dollar, how will β_1 and β_0 change?[1]

(c) You have been moved in the Tax Department and you have a boss who wants to collect money by rising taxes on cigarettes. Will you advice him to raise the taxes on cigarette?

---

[1](Assume that all the packs contain 20 cigarettes and that 1=$1.42.)