

SUNY-Stony Brook. Economics Department
Economics 323: Fall 2011
Professor Hugo Benítez-Silva

Problem Set 5: Suggested Solutions

1 True-False and Multiple Choice (Again, justify all your answers).

(15 points: 5 points each)

A) Adding an exogenous variable to a regression model always improves the R-bar squared. **True or False.**

False, as discussed in class the R-squared always improves but the R-bar squared might go down, since it takes into account that we are estimating one more coefficient.

B) Present oriented people discount future earnings _____ forward looking people

a)at the same rate as

b)at a higher rate than.

c)at a lower rate than.

d)at a higher or lower rate than.

C) Nina (age 19) would be more likely to obtain a college education if

a)her discount rate was lower

b)the supply of others with a college education was greater

c)she expected to retire at a younger age

d)she was older

2. Short Essay questions (Answer should not be much longer than a paragraph or two, think before you answer)

(20 points 10 points each)

2a) A few years ago, the New England Medical College inadvertently accepted more applicants than it could accommodate in its first year class. Not wanting to arbitrarily delay the entrance date of its accepted applicants, it offered them one year of free tuition if they would delay their medical studies by one year. Discuss the factors entering into the student's assessment of whether he or she should take this offer.

Nothing special here, you had to consider the discount rate, the opportunity costs, and whether the labor market could change much from one year to the next. Coincidentally, a lot of students took the offer from the Medical College.

2b) In many countries higher education is heavily subsidized by the government (that is, university students do not bear the full cost of their college education). While there are many good reasons for heavily subsidizing university education, there are also some dangers in it. Using human capital theory, explain what these dangers are.

We discussed this a bit in class, if the cost of education is low more people are going to attend, even those for whom the benefits are relatively low, but their costs are subsidized. This can lead to over education, tight labor markets and at the same time high unemployment rates, given the large supply of educated individuals and the low supply of other less skilled workers

3 Betty, you can call him Al.

(25 points)

Betty works in sales but is considering quitting work for two years to earn an MBA. Her current job pays \$40,000 per year (after taxes), but she could earn \$55,000 per year (after taxes) if she had her masters of business administration. Tuition is \$10,000 per year and the cost of an apartment near campus is equal to the \$10,000 per year she is currently paying. Betty's discount rate is 6% per year. She just turned 48 and she plans to retire when she turns 60, whether or not she gets her MBA. Based on this information, should she go to school to earn her MBA? Explain carefully.

The simplest way to do this was to do a NPV calculation, where the costs are the 10k of tuition plus the 40k of foregone earnings, and the benefits are the additional 15k Betty gets until she retires and after she finishes the MBA. This assumes that she cannot work during her studies, and that the psychic costs and benefits are basically negligible.

NPV (to the current period when she is 48) if you assume she starts to study when she is 48, and then works until the end of the year she turns 60: \$14,436

NPV (to the current period when she is 48) if you assume she starts to study when she is 49, and then works until the end of the year she turns 60: \$6,587

- 4. In my webpage you can download (right click and save) in Excel format (.xls) and Minitab format (.mtw) data on 85 countries in the late 1980s taken from the Human Development report of the United Nations. (I recommend you use Minitab to complete this exercise)**

There are three variables in the data set:

- (c1) Life Expectancy at birth
 - (c2) Income per capita in US\$
 - (c3) Index of access to health care (higher number means better access, 100 is the maximum)
- a) Using measures of central tendency and dispersion provide a good description of each of these variables.

Nothing special was expected here, you had to provide most of the measures we used in class, or a subset you consider appropriate. Ideally, you would compute the standard deviation of each of them and provide confidence intervals for the mean.

- b) Graph using Minitab or Excel each of these variables against one another. Describe what you conclude from these graphs. From the Excel file you can read the names of the countries to which these data correspond. Any outliers in terms of the relationship between income and access to health care?

Graphs hopefully showed that life expectancy seems to be positively related to both income per capita and the index of access to health care. Income per capita is again a tricky variable, similarly to what we saw in class when analyzing fertility, because most countries have rather low levels and then a few have quite large. This will make our life a bit more difficult.

There are quite a lot of countries with fairly low income per capita and good health care index, like Korea, or even Burundi, Chile or Costa Rica. Among the rich countries most of them have high health care indices, Hungary seems to be an outlier, but my feeling is that they made a mistake when coding the income per capita level. I would be tempted of dropping that observation altogether, it is suspicious.

c) Can you think of an economic model that relates life expectancy with these variables? Explain.

Here you had to use some economic intuition to argue that you expect life expectancy to be related to income per capita and/or access to health care, since both things affect how long people live.

d) Estimate the parameters of the following simple regression model

$$\text{Life Expectancy}_i = \beta_0 + \beta_1 \text{Income per capita}_i + u_i$$

Interpret the estimates of the coefficients of interest and the overall regression.

$\beta_0 = 57.6$ (life expectancy would be around this age even with zero income)

$\beta_1 = 0.001293$ (if income per capita increases by \$1000 you would see the life expectancy increase by about 1.3 years.

The R-squared is quite good about 0.5. We can explain about 50% of the variation in life expectancy.

If you drop Hungary from the sample then,

$\beta_0 = 57.45$ (life expectancy would be around this age even with zero income)

$\beta_1 = 0.001363$ (if income per capita increases by \$1000 you would see the life expectancy increase by about 1.36 years.

The R-squared goes up to 0.5151.

e) Estimate the parameters of this other simple regression model

$$\text{Life Expectancy}_i = \beta_0 + \beta_1 \text{Access to Health Care}_i + u_i$$

Interpret the estimates of the coefficients of interest and the overall regression.

These results have already dropped Hungary.

$\beta_0 = 35.88$: life expectancy would be around this age even with zero health care index. Notice this is much lower than with the previous model.

$\beta_1 = 0.3608$: if the health care index increases by 1 point you would see the life expectancy increase by about 0.36 years, or if the health care increases by 10 points the life expectancy goes up by 3.6 years.

R-Squared is around 0.7186, much higher than with the income per capita as regressor. More than 2/3 of the variation in life expectancies can be explained by the variation in health care access. This is quite high.

f) What is your opinion of policies promoted by the United Nations to increase access to health care in developing countries? Do you think the simple regression models are enough to make policy recommendations?

Taking the regression results at face value indicate that the policies by the U.N. are very appropriate. It is debatable whether this simple model is enough to make policy recommendations, after all we know many other things affect life expectancy, and for the moment we have assumed that only the health care matters. If we are to spend millions of dollars making health care more accessible we have to make sure it will work, it might be that there are some other problems, like transportation costs, that could make this particular investment a very poor one. In any case this initial evidence is quite compelling so we at least analyze the link further.

5 Education and Earnings revisited

(25 points, 5 points each section, aprox.)

We have been talking for some time now about a model that tries to explain the differences in labor earnings across individuals with an array of variables, the most important of which we have assumed is education. Few people would argue against the fact that there is a correlation between education and earnings, but as economists our objective is to quantify the strength of the relationship.

In this exercise you are going to use another data set, the NELS, you can find information about this data set at: <http://nces.ed.gov/surveys/nels88/>

This is a nationally representative sample of eighth-graders who were first surveyed in the spring of 1988. A sub-sample of these respondents were then resurveyed through four follow-ups in 1990, 1992, 1994, and 2000. On the questionnaire, students reported on a range of topics including: school, work, and home experiences; educational resources and support. For the three in-school waves of data collection (when most were eighth-graders, sophomores, or seniors), achievement tests in reading, social studies, mathematics and science were administered in addition to the student questionnaire.

You can download a sub-sample of this data, with more than 8,500 observations in Stata and Excel format from the class webpage.

- a) The survey followed individuals for a few years and recorded the wages of all the respondents that were working as of the year 1999 and asked them for their earnings. It also recorded their education and other characteristics. Provide a quick summary of the main variables of interest, especially income in 1999, and 1998, the education indicators, and other socio-economic characteristics. (Hint: notice that in this case we have indicators for the highest degree obtained, not the number of years of education)

Here you just had to provide some of the usual summary statistics, nothing special, except noticing that for the education variables you had to provide proportions of people with that degree.

- b) Using first the simple regression model we presented in class compute the premium for having obtained a bachelor degree for this cohort as of their 1999 jobs. What about if you use the 1998 income? (Hint: Remember to use the logarithmic transformation discussed in class to get the premium in percentage terms)

23% for 1999 and 16% for 1998.

- c) How is this premium affected when you run a multiple regression model where you take into account variables like gender, race, and marital status? What about the effect of having dependents? And owning a house? Interpret the values of the other coefficients in the multiple regression model.

If you run the multiple regression with all the coefficients mentioned the college premium does not move that much, it is still around 20%. The coefficients on white and married are not very significant, and being male has a large positive effect, and so does owning a house. Having dependents (children, elderly parents) has a negative effect. The interpretation of the values is always in percentage terms of increase or decrease in income.

- d) Use the R-bar square to decide between models that include more of these additional characteristics. How many variables do you think should the multiple regression model, have to do the best possible job in explaining the variation in income as of 1999?

Here the more complete model does the job, the R-bar squared always seemed to go up. You were supposed to try different models and compare the R-bar squared.

e) Almost 3,000 of the respondents have a bachelor degree, how many have a Ph.D?
What is the premium as of 1999 from having a Ph.D?

Only 49 people have a Ph.D. in a simple regression model the premium is a bit higher than 19%. If you are to compare it directly with the college premium you can put other dummies except for bachelor. In that case a Ph.D. seems to increase, on top of the college degree, income by 13% or so.

f) One of the great advantages of the NELS is that we have the test scores of the respondents as of the time they were 13. Using the math test scores in our multiple regression model, explain the effect on income and the college premium. Is it what you expected? What about the other test scores, like reading, science, etc? Do the same using the income as of 1998. Do you think these tests are a good measure of ability?

Once we put the tests in general the college premium goes down, the math test seems to be a positive indicator but reading is negative, and others are not very significant. The changes were to be expected since these are measures of ability, and we had discussed the possible ability bias, which would increase the college premium if we did not control for these measures. The results are similar with the 1998 income.