Instructions

Your 1st Assignment has to be send via Email to repasky@tu-berlin.de by **19.12.2016 at 24:00 latest!** Bear in mind that assignments handed in **too late will not be corrected.** This is a part of your exam, therefore, cheating (not allowed cooperation between groups) will result in a failing final grade. **Your PDF document cannot contain more than 1600 words.** After 1600 Words, I will stop reading!

Every assignment consists of:

1. A pdf document containing your answers as well as explanations. **The graphs and tables, to which you are referring, have to be included in the document as well.** Do not include any parts of your code in the pdf document. Your solution needs to be typed on the PC, not hand-written and scanned. You can use Word, LaTeX or any other text-producing software. (But your final document needs to be a pdf)
   (Hint: With Word you can use the built-in Formula-Editor or install Math-type (3rd party software featuring 30-days free trial version. The free version of the product (even after the 30 days) is still good enough.)
   (Hint 2: You can create a pdf documents out of any word document pretty easily using any "pdf-creator" software such as pdf24 for windows user.)
2. Well-documented R-script or a STATA-dofile, which replicates your results. No Log-File is necessary.
3. STATA-users have to use the version control defining the version of STATA to 11.2 (Just state "version 11.2" at the beginning of your do-file (naturally without the quotation marks)).
4. Every assignment needs to include a tittle page with your 1.) Full Names, 2.) Student ID's, 3.) Your Course of Study and University.

In sum, you have to supply only 2 documents (a pdf containing your answers and explanations and a R-document (or STATA do-file) containing your code and comments explaining your code).

Task 1 (Linear Probability Model):

a) Please estimate the linear probability model of the probability of having finished at least 2-years college (`some_college = 1`) on the following explanatory variables:
`hscath` (= 1 if catholic high school graduate),
`hsrural` (= 1 if high school rural),
`grades` (average grade on 13 point scale with 1 = highest),
second order polynomial in `faminc` (gross 1991 family income (in $1000)),
`famsize` (number of family members),
`asian` (= 1 if asian),
`hispan` (= 1 if hispanic),
`black` (= 1 if black),
`par_college` (=1 if at least one of the parents have a college education).

   i. State the t-value for the explanatory variable `par_college`. What exactly can you learn from that value? (Explain briefly but be specific!)

   ii. Please interpret properly the effect of the family income on a median individual in your sample.

   iii. Plot the effect of (a marginal increase in) family income (`faminc`) on the probability to finish at least a 2-year college for all values of `faminc` in your sample. How would a similar graph for the variable `grades` differ? (Briefly explain!)

   iv. One of the downsides of the LPM is that the predicted probabilities could be bigger than 1 and smaller than 0. Find out for how many observations in your sample the predicted probabilities are bigger than 1 and for how many they are smaller than 0.

   v. Please describe differences in observed characteristics between the subset of sample that have predicted probability bigger than one and the rest of the sample. (Concentrate on the 3 biggest differences in mean characteristics)

Task 2 (Logit, Probability effects and the Delta method):

a) Please re-estimate the model from the Task 1 using logit.

   i. Can you learn anything from the estimated coefficients? (Be specific!)

b) Please compute the discrete probability effect (DPE) of the Variable `par_college` on the probability of having finished at least a 2-years college. Please assume the **median** values for all other explanatory variables. (Do not use any preprograme packages (R) or command (STATA) to compute the DPE).

   i. Instead of the median values, you could have assumed the mean values for the explanatory variables. Do you see any advantage/disadvantage of the median values over the mean values? (Name and explain one advantage **or** disadvantage that you thing is the most important!)

   ii. Another possibility for computing the probability effect of the variable `par_college` is the marginal probability effect (MPE). Discuss which alternative (DPE or MPE) makes more sense in this particular example and why?

   iii. <u>True or false and explain</u>: The DPE will be always larger than MPE (for the same variable).

   iv. For you to be able to say anything about the population's DPE, you need to compute the standard error of the DPE computed in 2b). What method would you use to compute it and why? (Be very specific which method and why it is necessary in this example!)

   v. Use the method you specified in iv.) to compute the standard error of the discrete probability effect computed in 2b.) (You are not allowed to use any preprogrammed packages or commands beyond the Variances of the β's)

   <u>(If you are not able to compute the standard error, then assume any value for it and carry on with the Task!)</u>

   vi. Perform an appropriate statistical test to find out whether the DPE computed under 2b.) **is bigger than 4 percentage points**.

   - <u>Reason very specifically,</u> **why the test** you have chosen **is appropriate**! (i.e. Distributional assumptions, one-sided or two-sided version)
   - Interpret the result of the test properly!

   vii. Interpret the computed DPE properly!