INTERNATIONAL UNIVERSITY OF JAPAN
Public Management and Policy Analysis Program
Graduate School of International Relations

ACC6512 (Winter 2012)
Topics in Data Analysis: Panel Data Models

<u>Homework 1: Ordinary Least Squares and Hypothesis Test (100 points)</u>

Instructions: Read data description carefully. Prepare `.do` file to manipulate data and answer questions. Don't forget to use log command. Print out `.do` file and log file. Handwrite your answer on the print-out of the log file; you may use other papers, if really needed. Use the .05 significance level in homework assignment. This assignment is submitted to my mailbox (or office) by Sunday, Feb. 5th. If you are not comfortable reading OLS output, see Freund and Littell (2002) http://www.sonsoo.org/documents/Freund_Littell_2000_SAS_Regression.pdf

Question 1 (10 points) Get WHO health care attainment data from William Greene's Web page http://www.stern.nyu.edu/Ewgreene/Econometrics/WHO-data.xls. Import it into Stata and lowercase all variables names. Remove observations if their `groupti` is not equal to 5 in order to get a strongly balanced panel. Create a numeric variable `country` that has country codes. Keep OECD countries only.

Question 2 (10 points) Fit a linear regression model to explain a dependent variable, health care attainment (`comp`). You must include at least health expenditure and educational attainment (or their transformation) and the number of regressors in your model should be larger than 3 and smaller than 7 [4, 6]. Do not include year or time related variables. Your model should be statistically significant (at least health expenditure must be a significant regressor) so that your interpretation makes sense; Do not try to interpret a model whose parameter estimates are all zero. Explicitly state your model specification on the print-out of the log file (near regression part).

Question 3 (10 points) Circle SSE (Sum of squares of errors) on the ANOVA table on the print-out of the result. Show how it is calculated using Stata `.predict` command.

Question 4 (5 points) Circle df_{total}, df_{model}, and df_{error} on the ANOVA table. Show how they are calculated on the print-out.

Question 5 (5 points) Circle MSM (mean squares of model or regression) and MSE (mean squares of error) on the ANOVA table. Show how they are calculated on the print-out.

Question 6 (5 points) Circle F score on the ANOVA table. Show how they are calculated on the print-out.

Question 7 (5 points) Conduct a hypothesis test of overall fit of the model. Use the p-value approach. State the null hypothesis of the F-test clearly.

Question 8 (5 points) Circle R^2 and adjusted R^2 on the ANOVA table. Show how they are calculated on the print-out.

Question 9 (5 points) Circle SEE or Root MSE on the ANOVA table. Show how it is calculated on the print-out. How do you interpret SEE?

Question 10 (5 points) Circle parameter estimates and their standard errors on the print-out. After rung `.regress`, run `.matrix list e(b)` and `.matrix list e(V)` to get a vector parameter estimates and their variance-covariance matrix. Remember that Stata is casesensitive. Show how standard errors of health expenditure and educational attainment are calculated.

Question 11 (5 points) Circle t score and p-value of health expenditure on the print-out. Conduct the corresponding hypothesis test using the test statistic approach. Begin with stating the null hypothesis and do not skip any step.

Question 12 (5 points) Conduct the same hypothesis test running `.test hexp=0` (you do not repeat the same steps except for calculating the test statistic). Show how F score is equivalent to the t score in question 11.

Question 13 (5 points) Circle t score of educational attainment on the print-out. Conduct the corresponding hypothesis test using the confidence interval approach. Obviously, you need to form a 95percent confidence interval first.

Question 14 (5 points) Conduct the hypothesis test that both parameters of health expenditure and educational attainment are zero. Show your computation in this Wald test. Run, for example, `.test hexp=hc3=0` to make sure your computation. See the result of question 10. Use the p-value approach.

Question 15 (5 points) Circle t score and p-value of a regressor you chose on the print-out. Conduct the corresponding hypothesis test using the p-value approach.

Question 16 (10 points) Circle the parameter estimate of health expenditure. Interpret it at least two different ways. Do not forget to add *ceteris paribus* assumption.

End of homework assignment 1.