

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**

Homework 2: Least Squares Dummy Variables (100 points)

**Instructions:** Use the WHO health care attainment data that you prepared for homework assignment 1. This assignment needs to be submitted in the class on Thursday, Feb. 9<sup>th</sup>. Use the .05 significance level and the p-value approach for hypothesis test.

**Question 1 (5 points)** Fit your linear regression model used in homework 1. You may change specification under the same conditions (e.g., You must include at least health expenditure and educational attainment as regressors). Let us call this model a pooled OLS. Create five dummy variables for year 1993 through 1997 and name them as `year93`, `year94`, ... `year97`. Do not use `t93-t97` when creating `year93-year97`.

**Question 2 (50 points).** Least Squares Dummy Variables 1 (LSDV1)

Question 2.1 (**2 points**) Fit LSDV 1 using the time dummy variables you created in Question 1. Omit `year95` in order to use year 1995 as the reference point (baseline).

Question 2.2 (**5 points**) Conduct the hypothesis test (F-test) for overall fit of the model. Is there any big difference in the F-test between the pooled OLS and LSDV1?

Question 2.3 (**3 points**) Report SSE, SEE,  $df_{\text{error}}$ , and  $R^2$ . Compare them with corresponding statistics in the pooled OLS in Question 1. Did you find any big difference?

Question 2.4 (**2 points**) Report coefficients and their standard errors of health expenditure and educational attainment. Do they differ from those in the pooled OLS?

Question 2.5 (**5 points**) Report five regression equations. Show necessary computation.

Question 2.6 (**3 points**) Interpret the intercept substantively. What does that mean?

Question 2.7 (**3 points**) Draw regression lines on a plot. Do not use rulers; Handwrite will do.

Question 2.8 (**5 points**) Conduct the t test for parameter estimate of `year97`.

Question 2.9 (**2 points**) Interpret the coefficient of `year97` substantively.

Question 2.10 (**10 points**) Conduct the hypothesis test to see if all four dummy parameters used in the model are zero. Follow all steps for hypothesis test. You may run `.test year93=year94=year96=year97=0`. Interpret the result (Did inclusion of a set of dummy variables improve your model?)

Question 2.11 (**3 points**) Fit the LSDV1 using year 1993 as the reference point (i.e., dripping `year93` instead of `year95`). Comparing with LSDV1 in Question 2.1, was there any change in goodness-of-fit measures (e.g., SSE, SEE, F, and  $R^2$ )? How about parameter estimates of health expenditure and educational attainment?

Question 2.12 (**5 points**) Report regression equations of the model in Question 2.11. Show your calculation. Do they differ from those in Question 2.5? Interpret intercept and coefficient of `year97` substantively.

Question 2.13 (**2 points**) Run the model in Question 2.11 using the `.xi` command (NOT `.regress`). You should get the identical result.

### **Question 3 (20 points).** Least Squares Dummy Variables 2 (LSDV2)

Question 3.1 (**2 points**) Fit LSDV2 using `.regress`.

Question 3.2 (**3 points**) Report SSE, SEE,  $df_{\text{error}}$ ,  $R^2$ , and adjusted  $R^2$  and then compare them with those of LSDV1 in Question 2.1.

Question 3.3 (**2 points**) Report coefficients and their standard errors of health expenditure and educational attainment. Do they differ from those of LSDV1 in Question 2.1?

Question 3.4 (**5 points**) Report five regression equations. Show necessary computation.

Question 3.5 (**5 points**) Conduct the t test for parameter estimate of `year97`.

Question 3.6 (**3 points**) Interpret the coefficient of `year97` substantively.

### **Question 4 (25 points).** Least Squares Dummy Variables 3 (LSDV3)

Question 4.1 (**2 points**) Fit LSDV3 using `.cnsreg`.

Question 4.2 (**3 points**) Report SSE, SEE,  $df_{\text{error}}$ ,  $R^2$ , and adjusted  $R^2$  and then compare them with those of LSDV1.

Question 4.3 (**2 points**) Report coefficients and their standard errors of health expenditure and educational attainment. Do they differ from those of LSDV1?

Question 4.4 (**5 points**) Report five regression equations. Show necessary computation.

Question 4.5 (**5 points**) Report the intercept and show how it is calculated from the result of LSDV2.

Question 4.6 (**5 points**) Conduct the t test for parameter estimate of `year97`.

Question 4.7 (**3 points**) Interpret the coefficient of `year97` substantively.

End of homework assignment 2