

The Code

/\* This is a solution to the application problem on pages 49--50 of Greene's text \*/

/\* Read in the data \*/

**data** example1;

input person\_id education wage experience ability mother\_ed father\_ed siblings;

cards;

1 13 1.82 1 1 12 12 1

2 15 2.14 4 1.5 12 12 1

3 10 1.56 1 -0.36 12 12 1

4 12 1.85 1 0.26 12 10 4

5 15 2.41 2 0.3 12 12 1

6 15 1.83 2 0.44 12 16 2

7 15 1.78 3 0.91 12 12 1

8 13 2.12 4 0.51 12 15 2

9 13 1.95 2 0.86 12 12 2

10 11 2.19 5 0.26 12 12 2

11 12 2.44 1 1.82 16 17 2

12 13 2.41 4 -1.3 13 12 5

13 12 2.07 3 -0.63 12 12 4

14 12 2.2 6 -0.36 10 12 2

15 12 2.12 3 0.28 10 12 3

;

/\* The usual checks \*/

**proc** **contents** data=example1;

**run**;

**proc** **print** data=example1;

**run**;

/\* Part a: OLS of wage versus (constant, education, experience, ability) \*/

**proc** **reg** data=example1;

model wage=education experience ability;

**run**;

/\* Part b: OLS of wage versus all the variables (but for id, of course) \*/

**proc** **reg** data=example1;

model wage=education experience ability mother\_ed father\_ed siblings;

**run**;

/\* Part c: OLS of each of mother's education, father's education and siblings on the regressor set in

Part a. We save the residuals and compute their means \*/

/\* Note that by definition, the mean of the residuals is equal to 0 \*/

**proc** **reg** data=example1;

model mother\_ed=education experience ability;

output out=resid1 r=r1;

**run**;

**proc** **reg** data=example1;

model father\_ed=education experience ability;

output out=resid2 r=r2;

**run**;

**proc** **reg** data=example1;

model siblings=education experience ability;

output out=resid3 r=r3;

**run**;

**proc** **means** data=resid1;

var r1;

**proc** **means** data=resid2;

var r2;

**proc** **means** data=resid3;

var r3;

**run**;

/\* Part D: Compute R-Square for Part B and then same but without the intercept term \*/

/\* Note that the results from Part B already has the R-Square value \*/

/\* We re-run the analysis without the intercept term \*/

/\* Part b: OLS of wage versus all the variables (but for id, of course) \*/

/\* R Square and the F statistic values both increase--Why? \*/

**proc** **reg** data=example1;

model wage=education experience ability mother\_ed father\_ed siblings/noint;

**run**;

/\* Part E: Calculate the Adjusted R Square value. Note that this was done under Part B above \*/

/\* Part F: We need to regress wages on the regressor set from Part A and the three residual variables from Part C \*/

**data** temp1;

set resid1;

keep person\_id r1;

**run**;

**data** temp2;

set resid2;

keep person\_id r2;

**run**;

**data** temp3;

set resid3;

keep person\_id r3;

**run**;

**data** new\_set;

merge example1(in=a) temp1(in=b) temp2(in=c) temp3(in=d);

by person\_id;

if a and b and c and d;

**run**;

**proc** **reg** data=new\_set;

model wage=education experience ability r1 r2 r3;

**run**;