This assignment builds on Assignment #3. If you have not already done assignment #3, you must do it first in order to build on it for this assignment.

1. Present your regression estimates for the replication part of Assignment 3
   a. (That is, run regressions like those in Table 2 of Pendakur and Pendakur 2010, corresponding to regressions controlling for personal characteristics only in Vancouver in 2005 (2006 microdata are about 2005 earnings). Be sure that you get the dependent variable correct. Report the output for the visible minority and Aboriginal coefficients only.)
2. Endogeneity
   a. Control for field of study. Report the resulting regression estimates.
   b. What do you do with people who do not report a field of study? (Why don't they report a field of study?)
   c. What is the meaning of the new coefficients in your regression?
   d. Does inclusion of field of study change your estimates for visible minority or Aboriginal coefficients?
   e. Test whether or not visible minority or Aboriginal men or women have lower log earnings than white men or women, respectively.
   f. If it changes the coefficients, does this mean that the estimates Pendakur and Pendakur (2010) present are biased?
   g. Should you control for field of study?
3. Heteroskedasticity
   a. Is the variance of the log of earnings smaller different between Aboriginals, Visible Minorities and Whites?
   b. If so, does this affect the interpretation of the OLS reported standard errors? How so?
   c. Run the regression (including field of study) reporting White hetero-robust standard errors.
   d. Redo the hypothesis tests of 2e. Do the results of the tests change? Why or why not?
   e. If so, how might this induce inefficiency in the OLS estimator?
   f. Can you correct for this using weighted least squares? How does weighted least squares reduce the variance of the estimator?
   g. Estimate the model (including field of study) by weighted least squares. Report the resulting estimates.
   h. Are the estimates different?
   i. Do the hypothesis tests in 3d change? Why or why not?
4. Specification
   a. Instead of regressing log earnings on regressors, regress earnings on regressors (including field of study). Report the regression results.
   b. What is the difference in the interpretation of the coefficients in this regression versus that you reported in 2a?
   c. Are the coefficients different than in 2a? Are hypothesis tests different than in 2e?
   d. No, run the regression reporting White hetero-robust standard errors. Report the results. Do the results of the hypothesis tests change?
   e. Run regressions like those in Table 2 of Pendakur and Pendakur 2010, corresponding to regressions controlling for personal **and work** characteristics in Vancouver in 2005. Do not control for field of study. Use regular OLS standard errors (not White hetero-robust ones). Report the output for the visible minority and Aboriginal coefficients only.
   f. How do you interpret the difference between these coefficients and those you reported in 1a?
   g. Should you control for personal and work characteristics? Or just personal characteristics?