Problem Set 5 Hints

Exercise 1.

1. You are asked to run to two separate t tests. One t test which compares the mean `logno2` in the treatment areas in the before and after periods. A separate t test which compares the mean `logno2` in the control areas in the before and after periods. There are several ways to run these t tests (You should get very similar if not identical answers using all three methods; differences due to rounding):

- From material covering hypothesis testing for means: Calculate them by hand using the formulas previously discussed in lecture and notes.
- From the STATA tip noted in the problem set: Use the `ttest` command.
- From Handout #18 (Example from Kiel and McClain): Regress `logno2` on the after indicator for only those observations in the treatment areas; run an identical regression for only those observations in the control areas. We didn't explicitly discuss this in lecture but it's a useful trick.

2. To be consistent with Wooldridge's (4e) definition of the difference-in-differences estimator presented in equation 13.6, the parameter representing the difference-in-differences estimator in the difference-in-differences regression equation is given by the coefficient on the policy indicator (i.e. the interaction term). The question asks you to interpret this parameter; however, an extra credit point will be given for correctly interpreting all of the regression coefficients in the difference-in-differences regression equation.

When asked to 'perform the estimation for NO2' this just means that you need to use the data to estimate the regression equation you wrote down in the first part of this question. It is intended that you use `logno2` as the outcome variable so no need to complicate things (although it's okay if you un-log this and use NO2 as the question reads).

Exercise 2.

See the solutions to Daily Assignment #16 as well as the lecture notes on this material; I provide ample guidance.

Exercise 3.

(a) To make the post period clearer, I've revised the window for the birth data so that births are observed between 2003 and 2010; I also assume that the fracking site becomes active Jan. 1 2006. Thus, the post period is the year 2006 and after. Here is the revised summary:

"Hill uses a difference-in-differences research design to estimate the effect of proximity to active fracking sites on birthweight in Pennsylvania. The treatment group consists of all births between 2003 and 2010 whose mother's residence is within 2.5 kilometers of a site that eventually has active fracking. The control group consists of all births between 2003 and 2010 whose mother's residence is within 2.5 kilometers of a site that is eventually licensed to have fracking but does not become active during the study period. For simplicity, assume that that sites that eventually had fracking become active on Jan. 1, 2006."

(c) In order for an omitted variable to introduce bias it must be:

- _____ and
- _____

(e) I am making this a bonus question. Here is some discussion to chew upon: Many clinical drug trials are based on some biological mechanism, but not always. Do clinical drug trials tell us the cause of the effect produced by a drug or do they help us recover the effect of taking the drug despite imperfect knowledge of the cause?

Exercise 4.

- The key to all of these questions is to think about:
  - What types of unobserved factors does the difference-in-differences regression equation control for?
  - What types of unobserved factors does the diff-in-diff regression equation not control for?
- ***Interpreting bias away and towards zero (*** $\hat{\beta} = \beta + bias$ ***)***
  - If the true effect is postive ($\beta > 0$), then:
    - a negative bias would bias the estimated effect ($\hat{\beta}$) towards zero and
    - a postive bias would bias the estimated effect ($\hat{\beta}$) away from zero.
  - If the true effect is negative ($\beta < 0$), then:
    - a negative bias would bias the estimated effect ($\hat{\beta}$) away from zero and
    - positive bias would bias the estimated effect ($\hat{\beta}$) towards zero.