

## Data Exercise

CEGA DeCal on Global Poverty & Impact Evaluation

Clair Null, Nov. 17, 2008

You will need to use the following Excel formulas: average, stdev, sqrt, rand

To use one of these formulas, just put your cursor in a cell, type “=” and then the formula name (e.g. “average”), then “(“, click and drag to highlight the values you want to include when evaluating the formula, and finally close the parentheses with “)”. For example, to calculate the average of the numbers in cells A1 through A5 you would type “=average(A1:A5)”.

You can also type in an equation, such as “=A1+1.96\*A2/sqrt(300)”.

1. Calculate the effect of spring protection and chlorine distribution using the differences-in-differences method and the data on the “large sample” worksheets.
2. What do you learn from the pre-intervention data? Do you really need these data?
3. Do the post-intervention differences between the treatment and control groups for either of these water treatment technologies seem to be due to chance? (Hint: how do the 95% confidence intervals for treatment and control compare?)
4. Compare the difference-in-difference estimates of spring protection and chlorine distribution. Which intervention had a larger effect?
5. Redo steps above using a subsample of the data. Do this for both the spring protection and chlorine distribution data.
  - a. To take a random subsample, use the random number generator in Excel. Put your cursor in cell A3 and type “=rand()” and then click and drag down to the end of the data (cell A302).
  - b. Now you want to sort the random numbers you just generated. I was having trouble doing this with the formulas, so let’s convert to values first: highlight cells A3 to A302, copy them (edit menu → copy), and paste as values right back into those same cells (with A3 to A302 still highlighted and blinking to show that it’s been copied, edit menu → paste special → values and then the OK button). Make sure this worked – if you put your cursor in cell A3 it should now show a number and not the “=rand()” formula. Sort the data according to the random numbers: highlight cells A3 through G302, data menu → sort (column A, ascending is fine). Hit OK.
  - c. We’ll use just a quarter of the sample. Copy the data in columns C through G for which the random number in column A is less than 0.25 (this will be around 75 observations, but not exactly). Highlight C3 through G75 (or so), edit menu → copy, put your cursor in cell A3 on the corresponding “small sample” worksheet and paste in the data (edit menu → paste).
  - d. Calculate the effect of spring protection and chlorine distribution using the differences-in-differences method and the data you just transferred into the “small sample” worksheets. (Hint: You can copy cells with formulas as well as cells with data.)
  - e. Write your estimated program impacts on the black-board (for both spring protection and chlorine distribution based on the small sample results).
  - f. Do the post-intervention differences between the treatment and control groups for either of these water treatment technologies seem to be due to chance? (Hint: how do the 95% confidence intervals for treatment and control compare?)
  - g. How do your results from the small sample calculations compare to your results from the large sample calculations?
6. Summarize your findings. What two factors affect your confidence that the estimated program effects are not just due to chance?