

3. **STATA exercise:** Download the STATA dataset “InfantMortalityAirPollution.dta” from Courseworks and open it in Stata. This data set was used in a recently published policy paper by Michael Greenstone and Rema Hanna, “Environmental Regulations, Air and Water Pollution, and Infant Mortality in India”. They analyze the impact of environmental regulations, such as mandating catalytic converters, on the quality of the environment – catalytic converters are emissions control devices for automobiles converting toxic pollutants in exhaust gas to less toxic pollutants. They also examine how environmental regulations affect health status indicators, such as infant mortality rates.

The data set contains information for a few randomly selected cities over multiple years. The variable `c_IM` records the infant mortality rate (in 1000 births) in a given city in a given year. At the beginning of the period analyzed, no cities had mandated catalytic converters. The variable `neveradoptCAT` indicates whether a city has changed its environmental regulation policy within the period analyzed. The variable takes value 1 in all years for a city that has never introduced mandates for catalytic converters, and 0 in all years for a city that has changed its environmental regulation policy. Finally, the variable `catconverter` takes value 1 if a city in a given year has mandates for catalytic converters and 0 if a city in a given year does not have such mandates.

- (a) (6 points) You want to test whether mandating catalytic converters lowers infant mortality rate. Basically, you want to test the Null Hypothesis $\mu_1 - \mu_0 = 0$ against the Alternative Hypothesis $\mu_1 - \mu_0 < 0$ where μ_1 is the mean infant mortality rate when a city has mandates for catalytic converters and μ_0 is the mean infant mortality rate when a city does not. Since you do not observe population means directly, you will have to engage in statistical inference based on the information from your samples. Write down the appropriate Stata command for this particular hypothesis test and attach its output.
- (b) (5 points) Using the Stata output above, answer the following questions: What is the sample mean value for the infant mortality in a year in which a city mandates catalytic converters? What is the sample mean value for the infant mortality in a year in which a city does not mandate catalytic converters? What are the values of $SE_{\bar{X}_0}$ and $SE_{\bar{X}_1}$? How many degrees of freedom would we use for your test if the number of degrees of freedom were not calculated by Stata automatically?
- (c) (6 points) Using the Stata output in part a), write down all the steps for the test of significance in part a) at the 5% significance level. Start with the Null Hypothesis and follow all the steps we detailed in class. You do not have to calculate anything - just use the information Stata output provides.
- (d) (6 points) You find the hypothesis test results surprising. So you are considering re-doing the test, but excluding this time from your sample the cities that never changed their environmental regulation policies - that is, using only the cities that switched from no mandates to mandating catalytic converter. Write down the Stata command and attach its output. Interpret the result using a 5% significance level.
- (e) (2 points) Without running a test again, would you be able to reject the null hypothesis for the test above at a 1% significance level? Why or why not?

- (f) (6 points) Using Stata, test whether the mean infant mortality rate for cities who never adopted Mandated Catalytic Converters is larger than the mean infant mortality for cities who adopted the regulation, using a 1% significance level. Write down the Stata command and attach its output. Interpret the result.

4. **STATA exercise:** Using the same data set, “InfantMortalityAirPollution.dta” run a regression of infant deaths per 1000 births (`c_IM`) on `year`, a variable showing the year in which the information about a city was collected.

- (a) (2 points) Run the regression. Show the STATA command and its output.
- (b) (4 points) Interpret the intercept of the regression. Is this a logical estimate? Why or why not?
- (c) (4 points) Interpret the slope of the regression. Is this a logical estimate? Why or why not?