

Problem Set 4

Vinish Shrestha

10/18/2021

Data work:

In this data work we want to explore the relationship between mother's education and infant health more closely and issues surrounding such relationships.

Open the data set "NCHS_birthweight2000_sample.csv" located in the **gitpage**. Note that this is a csv file. You will need to save the file in a particular location in your computer and state the path before you open the file. For example, if you store your file in this path: /home/desktop, then use the command:

`data <- read.csv("/home/desktop/NCHS_birthweight2000_sample.csv")` to read the file.

1. Run a univariate regression of the form: $birthweight = \alpha + \beta mother_education + \epsilon$. Report the coefficient on β . Does this show a causal relationship between mother's education and infant health. **Note that you should delete all observations with mother's education of 99 (not reported values) before you proceed with this.**
2. Next, generate an indicator variable for low birthweight called "low_bwgt" if $birthweight \leq 2,500$ grams.
3. Calculate the proportion of infants with low birthweight by mother's education. For instance, what proportion of infants belonging to mothers with 12 years of schooling have low birthweight. Do this for all education values.

`mean(data$low_bwgt[data$dmeduc == 12])` .. this gives the proportion of infants with low birthweights for moms who have 12 years of schooling.

`mean(data$low_bwgt[data$dmeduc == 13])` .. this gives the proportion of infants with low birthweights for moms who have 12 years of schooling.

4. Using your calculations from 3 plot the relationship between mother's years of schooling and proportion of infants with low birthweight. *hint:* type `help(plot)` in the console.
5. Now construct an indicator that represents mothers with more than high school education; "aboveHS" by using the following hint. `data$aboveHS[data$dmeduc>12] <- 1`. Follow similar approach to assign the value "0" for mothers with high school or less than high school.
6. Now run the regression of the model specification: $low_birthweight = \alpha + \beta above_HighSchool + \epsilon$. Comment on the coefficient of β .
7. Next, generate an indicator of race: black vs. white. Then run the following specification: $race = \alpha + \gamma above_HighSchool + \epsilon$. Comment about the point estimate of γ .
8. Using your realizations from 7, comment on the case of omitted variable bias when estimating $low_birthweight = \alpha + \beta above_HighSchool + \epsilon$ in 6.