Homework 0

Econ 322: Econometrics

Spring 2019

Prof. Amanda Agan

**Upload to Sakai by Wednesday February 20 at 11:55pm (Note sakai won’t let me do 11:59pm so be aware it is by 11:55pm!)**

Your document should include responses to Part I and Part 2. The answers themselves should be attached to the Sakai assignment as *one* .pdf file (export to .pdf from whatever software you use). Make sure your name is at the top. You may work together in small groups of 2-3 if you like, but the write up must be your own. You may not turn in identical write-ups. Please write the names of your collaborates on your homework as well. Also include your .R script as a separate upload.

**Part 1:** “Is maternal smoking associated with infant birth weight?"

* Download the Birth Weight dataset and the codebook from Sakai. The codebook gives you information about the variables in the dataset, it is important to read it before interpreting your results.
* First set your working directory to the *folder* where your data is (Session->Set Working Directory->Choose directory)
* Read the data into RStudio, data<-(‘birthweight.csv’, head=T, sep=’,’)
* Create a new data object, data2, which removes birthweights corresponding to mothers with unknown smoking status
  + See the R Tutorial
  + You need to look at the codebook to figure out which observations to eliminate
* Investigate whether maternal smoking is associated with infant birth weight.
* What would the null hypothesis for this question be? The alternative hypothesis?
* Create a table that shows:
  + - the mean birth weight of babies born to mothers that smoke and do not
    - the standard errors of those means,
    - the 95% CI of those means
    - the difference in these means,
    - the 95% CI of this difference
    - indicate the p-value for the difference via stars on the difference (with \*\*\* p<0.01, \*\* P<0.05 and \* p<0.1)
  + See the example table uploaded with this assignment – do not just cut and paste the exact R result, instead you must create your own “professional” table that reports the numbers that R gives you! See example Table template attached as part of this assignment.
  + The R Tutorial has instructions on how to do these things
  + You should also google around for R help!

* Do you reject or fail to reject the null hypothesis (and if you can reject, at what level can you reject)? How do you know? What does this mean?
* Create a write up of your analysis. The write up should include: A statement of the question being answered, table you created, and a summary of what you learn from this investigation which includes answers to the above questions listed questions.
* Also include a few sentences on whether you think this effect is causal? Why or why not?
* Your table should be “publication quality” with appropriate titles and labels and table notes

**Programming Resources**:

* <https://www.r-bloggers.com/how-to-learn-r-2/> Tutorials for learning R
* <https://stats.idre.ucla.edu/r/seminars/intro/> Introduction to R Seminar

**Additional Notes:**

* Your code should be saved in a .R file as good practice. This code should have comments at the top that include your name, and that this is code for Econ 322 Homework 0. Please turn in this .R file along with your homework

**Part 2:** Coming up with statistical questions of interest

* Download the Current Population Survey (CPS) data and codebook from Sakai. This dataset is a random sample of individuals from the 2012 CPS.
* First write a sentence or two describing the CPS (google CPS data to learn about it, be careful about plagiarism)
* Read the data into RStudio. What is the mean number of children under the age of 5? What percent of the respondents are male?
* What interesting statistical questions could you answer using the data provided? Focus on questions that deal with the relationship between two variables. Come up with at least 3 examples. You can load the data into Stata to look at it and you can also look at the codebook to get an idea of what variables are available.
* Choose one question: what is the null hypothesis for testing this question? The alternative?
* You do not actually have to do any data analysis for this section, just include the write up for the questions above