Econ 103 HW1 Part 2: Coding Part
Instructor: Lucas Zhang
2023 Fall

- This homework is due on **Oct 27th 11:59PM Pacific/LA Time** on course website. The late homework will not be accepted as the solution will be shown after that time.
- There will be designated sections on the course website for you to submit your homework.
- You are **encouraged to collaborate** with your classmates, but you **must submit your own individual copy**.
- You don't need to type your written response. You can take a picture of your handwriting, with questions' clearly numbered.
- Submit your code and your written response in `.pdf` format. You can either take a screenshots of your code or export the jupyter notebook as pdf.

Birthweight and Smoking

In the first empirical study, we are going to explore the relationship between infant's birth weight and whether the mother is a smoker.

1. The data `bw_smoking.csv` and the description of the data `Birthweight_Smoking_Description.pdf` are posted in the same directory as this homework on the course website. Read the description carefully.
   1.1 What is the description of the variable `birthweight`?
   1.2 What is the description of the variable `smoker`?
2. Download the data `bw_smoking.csv` from the course website, and upload it to the jupyter hub like we did in the lab.
3. Create a new notebook (or use the existing one).
4. Import the packages like we did in the lab. (Refer to `103_all_codes.ipynb` section 1)
5. Load the data `bw_smoking.csv` (Refer to `103_all_codes.ipynb` section 2)
6. What is the average birth weight for the entire sample? What is the average infants' birth weights for mothers who smoke? What is the average infants' birth weights for mothers who don't smoke?
   - To find the birth weights for mothers who smoke, you can create a new dataset consists of only mothers who smoke, and then find the mean using that dataset. Similarly for mothers who don't smoke
   - Refer to section 3 and section 4.1 of `103_all_codes.ipynb`.
7. Create a histogram of `smoker`. Refer to section 4.2 of `103_all_codes.ipynb`
8. Create a histogram of `birthweight`. Refer to section 4.2 of `103_all_codes.ipynb`

9. Create a scatter plot of `birthweight` (on the y-axis) and `smoker` (on the x-axis). Clearly label and title your graph. Refer to section 4.3 of `103_all_codes.ipynb`

10. Run a regression of `birthweight` on the binary variable `smoker` using the full dataset. Refer to section 5.1 of `103_all_codes.ipynb`

11. Create a scatter plot of `birthweight` (on the y-axis) and `smoker` (on the x-axis) with the regression line. Refer to section 5.3 of `103_all_codes.ipynb`

12. Interpret the coefficients from the regression. Using the results from the regression table, formulate and test the null hypothesis that there is no effect of mother's smoking on infant's birth weight. (You can either use t-statistic or using p-value)

13. From the table, what is the 95% confidence interval for the effect of smoking on infant's birth weight?

14. Create a scatter plot of residuals from the regression (on y-axis) and the regressor `smoker` (on x-axis). What are the definitions of homoskedasticity and heteroskedasticity? Are the residuals from this regression homoskedastic? Why or why not? Refer to section 5.4 of `103_all_codes.ipynb`

15. Does the OLS assumption, that the regressor `smoker` and the residual are uncorrelated, hold? Can you think of some other economic variables that maybe correlated with whether the mother smokes and with the infant's birth weight?