

## Homework 4: due Friday March 3

*Instructor: Donlapark Ponnoprat*

**1. Instrumental variable.** We will study the effect of education on the wife's hourly wage using 1975 labor supply data [Mroz, 1987], where each row is the data of a household. The data, which can be found in `wage75.csv`, has the following attributes:

Name	Description
<code>wage</code>	Wife's average hourly wage, in 1975 dollars.
<code>education</code>	Wife's education in years.
<code>experience</code>	Actual years of wife's previous labor market experience.
<code>feducation</code>	Wife's father's educational in years.
<code>fincome</code>	Family income, in 1975 dollars.

Since the education is not randomly assigned, we cannot estimate the effect of the education on the hourly wage directly. Fortunately, it is possible to estimate the effect (on the compliers) using the method of instrumental variables.

1. Only select the rows of data whose wife's hourly wage is greater than 0.
2. Choose one of `experience`, `feducation`, `fincome` to be the instrument, and explain why you think this variable would be a good instrument.
3. Make a scatter plot between the variable that you chose and the education. You might want to add small noises to one of the variables to see the relationship between the two variables more clearly.
4. Describe the compliers, defiers, never-takers and always-takers in this context.
5. For each of the four assumptions of instrumental variables, explain why your choice of the instrument might satisfy this assumption.
6. Use two-stage least squares to estimate the average effect, over the compliant households, of `education` on `log(wage)`. Report the estimate and the standard error.

**2. Regression discontinuities.** In this problem, we will estimate the effect of drinking on death by suicide using the mortality data as reported by National Center for Health Statistics [Carpenter and Dobkin, 2009]. The data, stored in `drinking09.csv`, shows the mortality rate for each age group. We will use the following two variables:

- `agecell`: the average age in each group
- `suicide`: the suicide rate per 100,000 individuals

In the US, the minimal drinking age is 21, so we may use 21 as the cutoff point for the regression discontinuity design; individuals under age 21 are in the control group, and those of age 21 and over are in the treatment group.

1. After loading the data, create a new dummy variable called `islegal` to indicate whether the age group is less than or not less than 21.
2. Plot the suicide rate by the age. Do you see the discontinuity at age 21?
3. We will perform regression discontinuity with `islegal` as the forcing variable. The ignorability assumption is automatically satisfied in this design, but is the continuity assumption also satisfied? Please explain clearly.
4. Fit a linear regression and report the estimate of the effect of drinking on the suicide mortality rate. A couple of notes:
  - (a) Adjust the forcing variable first.
  - (b) Only regress on a small interval around the cutoff point—you can choose your own size of the interval.
5. What is the difference between *average treatment effect* (ATE) and local average treatment effect (LATE)? Is the estimate that you just obtained an estimate of ATE or LATE?

## References

- [Carpenter and Dobkin, 2009] Carpenter, C. and Dobkin, C. (2009). The effect of alcohol consumption on mortality: Regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, 1(1):164–182.
- [Mroz, 1987] Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4):765.