PhD Econometrics (ECON50580)
Problem Set 4: RDD

This problem set will be graded. Rules:

- You have to work in groups of 3-5 students
- Submit your solutions in 1 pdf
- The code should be in the appendix
- Results should be presented graphically or in tables. No screenshots from statistical software!
- Use soft-coding in your code.
- For empirical exercises, show evidence that you used version control
- Submit via Brighspace

Submission deadline: Wednesday, March 24

A fundamental question in labor economics is whether a longer duration of unemployment benefits improve the quality of the job after finding employment. However, providing a causal answer to this question is difficult because unemployment duration is correlated with individual characteristics that may also affect job quality. In this exercise, we use a regression discontinuity design for causal identification. Under regular circumstances, an unemployed person receives unemployment benefits for 30 weeks. However, the benefit duration increases to 39 weeks if the person is at least 40 years old when he/she becomes unemployed.

The dataset *data_ps3.dta* is a sample of workers who have been laid off and subsequently found employment. The sample contains the following variables:

age	age at layoff
nonemp	non-employment duration in weeks
jobfind	a dummy that equals one if a person had a new job after 39 weeks
lwage0	log monthly wage in previous job
lwage1	log monthly wage in new job

1) Explain intuitively how this discontinuity in benefit duration can be exploited to estimate the causal effect of benefit duration on wages after re-employment.

2) Write down an estimating equation for a sharp regression discontinuity design whereby you control for the running variable with a second-order polynomial that is allowed to differ above and below the discontinuity. State and explain the identifying assumption that is necessary to interpret your coefficient of interest as causal.

3) Carry out two tests for the validity of the RD design: 1) plot the density of age at layoff; 2) plot the log previous wage against the age at layoff. For each test, produce a scatterplot with bin size 4 months (i.e. each point summarizes the average value on the vertical axis for workers whose age at layoff falls within that bin). For easier visual inspection, the plot should contain a vertical line at the discontinuity and lines for the second-order polynomials above and below the cutoff. What do those graphs tell us about the validity of the RD design?

4) Produce the main results graphically. We focus here on three outcomes, namely nonemployment duration, the probability of finding a job after 39 weeks and log wages in the new job. Use the same binned scatters as in exercise 3). Interpret your results.

5) Focusing on the same three outcomes, report the coefficient of interest (i.e. the coefficient of a dummy for being above or below the discontinuity) of the regression outlined in exercise 2). Produce a regression table with five panels

- The reduced-form effect at the discontinuity in the full sample using the regression from exercise 2).
- The reduced-form effect at the discontinuity with a bandwidth of ± 5 years in age at layoff.
- The reduced-form effect at the discontinuity in the full sample using a linear control for the running variable, allowing for different slopes above and below the discontinuity.
- The reduced-form effect at the discontinuity in the full sample, controlling for the running variable with a fourth-order polynomial and, allowing for different parameters above and below the discontinuity
- The reduced-form effect at the discontinuity based on the optimal bandwidth computed based on the producedure by **?**. You can use their Stata/R package *rdrobust*. The package offers multiple procedures to calculate the bandwith. Use the default procedure. Report the estimated bandwidth along with each coefficient.

Interpret and discuss the differences between the panels.