REGRESSION DISCONTINUITY DESIGN REFRESHER (CI in NLP course, SoSe 2022)

Suppose in a university system where students are automatically allocated to mandatory tutorial attendance in the second year if his/her/their first-year GPA is below 7 (out of 10, 10 is highest). Please do the following tasks:

1. Clear any data that is currently in Stata. Then set the number of observations to be 10000.
2. Draw 10000 values from a uniform distribution that is defined over the interval (0,10). Call this variable `Running`.
3. Generate a dummy variable that equals 1 if `Running`$\geq$ 7. Call this variable `D`.
4. Sort the data on the basis of `Running`. The first observation should have the lowest value for this variable. The last observation should have the highest.
5. Run the command `twoway line D Running`. Your friend Bob says that this means that the context is ideal for running a Sharp RD. Do you agree? Explain.
6. Run the command `preserve`. Drop all observations where `Running` is smaller than 5.5 or bigger than 8.5.
7. Generate a variable that equals `-3 + Running`. Call it `Grades0`. Generate a variable that equals `-1 + Running`. Call it `Grades1`. Plot `Grades0` and `Grades1` against `Running` on the same graph. Identify the potential outcomes for the persons who have a first-year GPA of 8. Which outcome will you observe in your data? Identify the potential outcomes for the persons who have a first-year GPA of 6. Which outcome will you observe in your data?
8. Run the following sequence of commands:

```
generate Grades2 = 4.3 + 2/(1 + exp(-20*(Running-7)))
replace Grades2 = Grades1 if _n>=1593
replace Grades2 = Running - 2.44 if _n<=1209
twoway line Grades2 Running, lcolor(gs11) xline(7) ///
xtitle("First-Year GPA") ytitle("Second-Year GPA") legend(off)
```

   (a) What regression specification would you use to estimate the causal effect of the assignment variable at 7? Why?

9. Run the following sequence of commands:

```
generate Grades3 = Grades0
replace Grades3 = 2 + (2/7)*Running if Running>=7
replace Grades3 = -13.5 + 2.5*Running if Running<7
```

```
generate Grades4 = Grades1
replace Grades4 = 4 + (2/7)*Running if Running>=7
replace Grades4 = -11.5 + 2.5*Running if Running<7
twoway line Grades3 Grades4 Running, xline(7) xlabel(5.5 6 6.5 7 7.5 8 8.5) ///
xtitle("First-Year GPA") ytitle("Second-Year GPA") legend(off)///
lpattern("dash" "dash")
```

(a) What regression specification would you use to estimate the causal effect of the assignment variable at 7? Why?

10. Run the following sequence of commands:

```
restore
generate NewAssignment = .3*sqrt(Running)
replace NewAssignment = .3*sqrt(Running) - .4 if Running<7
replace NewAssignment = .02 if Running<2
keep if (Running>=4&Running<=10)
twoway line NewAssignment Running, xtitle("First-Year GPA") ///
ytitle("Mandatory Attendance (Yes=1, No=0)") ylabel(0 1) ysc(titlegap(2)) ///
xlabel(4 5 6 7 8 9 10) xsc(titlegap(2)) ///
lpattern(dash)
```

(a) Your friend Bob says that this means that the context is ideal for running a Sharp RD. Do you agree? Explain.