Homework Assignment 6

R. Justin Ratcliffe - 12284830

Due: 9/6/2020

```
library(ggplot2)
library(dplyr)
```

Quantitative Analysis Questions

Question 1 (1 point)

No, this would not be a good running variable. We can see that the density of observations before and after the cutoff is disproportionate. With most of the observations occurring before the cutoff, we should be worried about manipulation and whether randomization holds.

Question 2 (1 point)

1.

This is not a fuzzy regression discontinuity design. The cutoff is clearly defined as less or more than 50% of the votes, and we can expect that the cutoff will be strictly enforced. The running variable is the share of votes.

2.

This is a fuzzy regression discontinuity design. While there is a stated cutoff of a 2.7 GPA, the fact that not every student will attend, we cannot expect that the cutoff will be strictly enforced. The running variable is the GPA, and the cutoff is a GPA of 2.7.

Question 3 (2 points)

Question 3(a) - (1 point)

Yes, rainfall would be a valid instrumental variable to use to get rid of bias. Rainfall is correlated with X, crop yields, but is uncorrelated with u, which includes the omitted variable, valuation of education by parents.

Question 3(b) - (1 point)

Step 1: First, we need to run the regression for crop yield X_1, and rainfall as Z_1, where $X_1 = \pi_0 + \pi_1 Z_1 + u$. We will use this regression to find $\hat{X}_1$. If there are multiple endogenous variables, we repeat this step for all value of X_i.

Step 2: Using the $\hat{X}_1$ obtained in stage one, we run the regression $Y = \beta_0 + \beta_1 X_1 + u$ to obtain the two-stage least squares. β_1 is calculated by dividing the Cov(Y, Z) by Cov(X_1, Z).

Question 4 (1 point)

Instrumental variable analysis, by design, examines the relationship between X and Y using the variation in X caused by Z. As a result, LATE provides a consistent estimate only for the subsample of the data that is related to Z, if the subsample is not representative of the relationship between X and Y. LATE will apply to the whole sample if all units in the sample have the same treatment effect. This means that $\beta_i = \beta$ for all values of *i*.

Programming in R Questions

Answers to questions 5 – 8 can be found in the submitted R script.

We would like to look at the spatial distribution of child poverty density rates in Chicago's neighborhoods.

For this we will use the `ComArea_ACS14_f.shp` provided in the homework folder. You will need to load the `tidyverse`, `sf` and `tmap` packages for this assignment

```
library(sf)
library(tmap)
library(tidyverse)
```

Question 5 - (1 point)

Load the Chicago neighborhoods data in the `ComArea_ACS14_f.shp` file and plot the basic geometry of the Chicago neighborhoods

Question 6 - (1 point)

Now create a new variable for the density of child poverty in each neighborhood (as of 2014). The child poverty information is stored in the `ChldPov14` variable. For example to calculate the population density based on population count in the neighborhood, we can do

```
chicago.dat <- chicago.dat %>% mutate(pop.density = Pop2014/shape_area)
```

Question 7 - (1 point)

After creating the variable for child poverty density, plot it using `sf` package and `tmap` package.

Question 8 - (2 points)

We would like to add another dataset on grocery stores in Chicago's neighborhoods. This information is `point` information.

```
# Get the point data from Chicago grocery stores location
groceries <- st_read("groceries.shp")

## Reading layer `groceries' from data source
`/Users/arvindilamaran/Documents/UCHICAGO/SUMMER2020/DPSS/hw6-
r/groceries.shp' using driver `ESRI Shapefile'
## Simple feature collection with 149 features and 14 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 1124188 ymin: 1826196 xmax: 1201803 ymax: 1950151
## projected CRS:  Transverse_Mercator

# Get the CRS (coordinate reference system) of the groceries point data
groceries_crs <- st_crs(groceries)
```

Question 8(a) - (1 point)

Now ensure that the same CRS system is used by the Chicago neighborhood's data

Question 8(b) - (1 point)

Now plot the child poverty density and grocery store location on the same map