SMI 606 (Intro. to Quant. Data Analysis)
Dr. Hartman

Assessment 2 (~500 words)

To help you complete this assessment, you may use resources (e.g., books, lecture notes, online guides, etc.) and/or work with other students in the class, but each person must hand in their own work (i.e., you are not allowed to copy text from someone else—that is plagiarism!).

R Markdown
You should upload your R Markdown file directly to MOLE, which will provide me with your R script, results, and any notes that you include to answer each question.

To create an R Markdown file, you should install the `rmarkdown` package via 'pacman' from the command line:

```
> pacman::p_load(rmarkdown)
```

And then render your R script from the command line (NOT from the R script you wish to run!):

```
> rmarkdown::render("your R script file location and name here.R")
```

Make sure you comment out (place a # in front of) any lines that will cause errors for the markdown process (e.g., 'setwd()', 'file.choose()', 'rmarkdown::render()', comments, etc.).

You should start your R script with the following header:

```
#' ---
#' title: "SMI 606 Assessment 2"
#' author: "Anonymous"
#' date: "`r format(Sys.Date())`"
#' ---
```

Remember, that at its most basic form, R Markdown is just an R script in which comments # are replaced by #' as in the following:

```
#' This is an example of a heading comment using the R Markdown formatting.
```

Mexican Elections and the Effect of the Conditional Cash Transfer Program
In this assessment, you'll be analyzing data from a study about the electoral impact of *Progresa*, Mexico's conditional cash transfer (CCT) programme. CCT programmes aim to reduce poverty by making welfare conditional upon the receivers' actions (i.e., the government only transfers money to individuals who meet specific criteria). For example, the CCT provided cash payments to families in exchange for regular school attendance, health clinic visits, and nutrition support. Cash payments were made from the government directly to families to decrease overhead and corruption. Eligible villages were randomly assigned to receive the program either 21 or 6 months before the 2000 Mexican presidential election. You can think of the treatment as villages which were assigned to the CCT 21 months before the election because they would have sufficient time to implement the programme.

The data for this assessment are available as the CSV file `progresa.csv` (as part of the QSS datasets on MOLE). Each observation in the data represents a precinct, and for each precinct the file contains information about its treatment status, the outcomes of interest, socioeconomic indicators, and other precinct characteristics (see below for more details about the variables included in the dataset). The sample contains precincts that had *at most* one participating village in the evaluation.

Names and Descriptions of Variables in the CCT Program (*Progresa*) Data

Name	**Description**
`treatment`	1 = treatment (precinct contains a village that received CCT at 21 months); 0 = control (village received CCT at 6 months)
`pri2000s`	PRI votes in the 2000 election as a share of precinct population above 18
`pri2000v`	Official PRI vote share in the 2000 election
`t2000`	Turnout in the 2000 election as a share of precinct population above 18
`t2000r`	Official turnout in the 2000 election
`pri1994`	Total PRI votes in the 1994 presidential election
`pan1994`	Total PAN votes in the 1994 presidential election
`prd1994`	Total PRD votes in the 1994 presidential election
`pri1994s`	Total PRI votes in the 1994 election as a share of precinct population above 18
`pan1994s`	Total PAN votes in the 1994 election as a share of precinct population above 18
`prd1994s`	Total PRD votes in the 1994 election as a share of precinct population above 18
`pri1994v`	Official PRI vote share in the 1994 election
`pan1994v`	Official PAN vote share in the 1994 election
`prd1994v`	Official PRD vote share in the 1994 election
`t1994`	Turnout in the 1994 election as a share of precinct population above 18
`t1994r`	Official turnout in the 1994 election
`votos1994`	Total votes cast in the 1994 presidential election
`avgpoverty`	Precinct Avg of Village Poverty Index
`pobtot1994`	Total Population in the precinct
`villages`	Number of villages in the precinct

Imagine that you've been hired as a consultant by the PRI-controlled Mexican government to determine the electoral impact of the CCT programme. In particular, party leaders want to know whether exposure to the CCT increased turnout and vote share for the incumbent party. They provided you the `progresa.csv` dataset and asked you to answer the following questions.

1. To begin, which measures of 1) voter turnout and 2) support for the PRI (in 2000) would you use for your analyses (list the specific variable names)? Explain your reasoning and discuss any statistical analyses that you conducted to inform your decision. Do the different versions of turnout and party support measure the same thing? Discuss.

2. The government believes that poor voters should have turned out in higher numbers in the 2000 election and cast their ballot for the PRI. Estimate the sample average treatment effects (SATEs) of exposure to the CCT using a difference-in-means estimator. Did the CCT programme increase voter turnout and PRI vote share in the 2000 election? Discuss your results.

3. Next, using a linear regression approach, estimate the same sample average treatment effects by regressing each outcome variable—turnout and vote share—on the treatment dummy variable (i.e., you will have two separate regression models). Interpret your results and compare them to your findings in Question 2.

4. The dataset also includes results from the prior election. Is there any evidence that poorer voters would support the PRI (and hence one of the reasons for introducing *Progresa*)? Examine whether precinct poverty is associated with support for any of the three major political parties in the 1994 election (check the correlations). Now regress each party's vote share on average poverty levels (rescaled from 0 to 1). Interpret the coefficient for average poverty for each of the three regression models. Do precincts with higher levels of poverty tend to vote for any particular party? Discuss.

5. Do you see any issues with the 1994 voter turnout measures? Discuss how you investigated this question and how it might impact your earlier analyses.