Problem Statement – DID Analysis

Part 1

The statistical analysis for this assignment should be done using either R or Python or SAS or Stata or SPSS programming languages/software packages.

Objective:

The goal of this assignment is to apply the Differences-in-Differences (DD) analysis using a dataset detailing bank operations from two Federal Reserve Districts before and after a significant policy change during the Great Depression.

This analysis should be done using either R or Python programming languages. If you plan to use any other programming language, do reach out to the professor.

Dataset:

The dataset for this Assignment is from You will analyze a dataset that includes information on the number of banks in the Sixth and Eighth Federal Reserve Districts, before and after the Monetary Policy.

[Interpret the results from each step]

Data Cleaning and Manipulation:

1. Explore the variables included in the dataset.
2. Load the data into your environment and perform any necessary cleaning steps.
3. Calculate and report the descriptive statistics for the number of banks open before and after the policy intervention in both districts. (consider policy intervention happened in November of 1930)
4. Create a line plot visualizing the number of banks over time for both districts, marking the policy intervention point.

Analysis:

1. Implement a DD estimation to calculate the effect of the monetary policy on the number of banks remaining open. (Refer DD PPT Slide 8)
2. Interpret the results from the above estimate.

Part 2

Objective and Dataset: Propensity Score Matching

The primary goal is to determine whether having a minimum requirement (such as a minimum number of committed buyers) for Groupon deals impacts the outcomes of these deals.

Specifically, the outcomes of interest are revenue, quantity sold, and Facebook likes received. To assess this, deals are divided into two groups:

the control group (deals without the minimal requirement) and

the treatment group (deals with minimal requirement).

Propensity score matching is used to analyze the effect of this minimum requirement.

Analysis:

Explore the variables included in the dataset.

1. Load the data into your environment and perform any necessary cleaning steps.
2. Explain what a propensity score is and how it is used in the context of this Groupon dataset. Why do we calculate propensity scores before performing matching?
3. Visualize the distribution between treatment and control groups.
4. Calculate a propensity score using a regression model (figure out which model to use).
5. Which variables in the above regression model were significant in predicting whether a Groupon deal belonged to the treatment group? Why are these variables important?
6. Conduct two-sample t-tests for 'revenue' and 'fb_likes', what conclusions can you draw? How do you interpret the p-values and confidence intervals in these tests?
7. What issues arose when 'min_req' was included in the logistic regression model for calculating propensity scores? Why did this make matching more difficult?

Deliverable: A Jupyter Notebook or R markdown implementing the analysis along with the interpretations and explanations required. If using STATA or SAS or SPSS, code and the output files are required.