Statistics Problem Set 1

DS 100

Professor Marc Ratkovic

November 16, 2023

Ahmed Salem (1981559), Jacob Chen (1980440), Tarlan Sultanov(1989479)

10.1 Questions

1. The unit of treatment assignment is the name of the email sender to the different state legislators. This allows the researchers to assess whether the state legislators' responsiveness changes because of racial discrimination.

2.

- Unit-Level Causal Effect under No Partisanship Signal:

  The unit-level causal effect is the difference in the likelihood of response between the DeShawn and Jake conditions when there is no indication of partisan preference.
- Unit-Level Causal Effect under a Republican Signal:

  The unit-level causal effect is the difference in the likelihood of response between the DeShawn and Jake conditions when a Republican partisan signal is present.
- Unit-Level Causal Effect under a Democratic Signal:

  The unit-level causal effect is the difference in the likelihood of response between the DeShawn and Jake conditions when a Democratic partisan signal is present.

3. **No Interference Assumption:**

   The "no interference assumption" in SUTVA states that the treatment assigned to one unit (e.g., legislator) does not affect the outcomes of other units.

   Extent of Satisfaction and Potential Violations:

   Extent: The assumption is satisfied if each legislator's response to an alias is independent of other legislators' responses.

   Violations: It would be violated if legislators' responses are influenced by the treatment assignment (alias) and subsequent responses of other legislators. This could happen through social influence, network effects, or shared information among legislators.

4.

```
# A tibble: 6 × 5
  treat_group        mean_leg_republican mean_leg_black mean_leg_latino mean_leg_white
  <fct>                            <dbl>          <dbl>           <dbl>          <dbl>
1 Jake, no party                   0.448         0.0690          0.0259          0.885
2 DeShawn Republican               0.447         0.0728          0.0247          0.884
3 Jake Republican                  0.446         0.0793          0.0329          0.862
4 DeShawn Democrat                 0.445         0.0616          0.0283          0.897
5 Jake Democrat                    0.447         0.0676          0.0375          0.876
6 DeShawn, no party                0.447         0.0806          0.0248          0.867
>
```

   the mean values for the pre-treatment covariates are relatively close across different treatment conditions. For example, the differences in means for mean_leg_republican, mean_leg_black, mean_leg_latino, and mean_leg_white are not drastically different between treatment conditions.

   The balance in pre-treatment covariates is important for establishing the internal validity of the results because it helps ensure that any observed treatment effects can be attributed to the treatment itself rather than systematic differences in covariates between treatment

groups. If there is a substantial covariate imbalance, it could introduce confounding, making it challenging to isolate the causal effect of the treatment.

5. The difference-in-means estimator involves taking the mean of the outcomes for treated units (Jake) minus the mean of the outcomes for untreated units (DeShawn) within each partisanship condition.

$$SATE_{\text{Republican}} = \frac{1}{N_{\text{Republican}}} \sum_{i=1}^{N_{\text{Republican}}} (Y_{\text{Jake},i} - Y_{\text{DeShawn},i})$$

$$SATE_{\text{Democrat}} = \frac{1}{N_{\text{Democrat}}} \sum_{i=1}^{N_{\text{Democrat}}} (Y_{\text{Jake},i} - Y_{\text{DeShawn},i})$$

These formulas represent the average treatment effect for Republicans and Democrats, respectively, by taking the mean of the differences in outcomes between the treated (Jake) and untreated (DeShawn) units within each partisanship condition.

6. We want to prove that these estimators are unbiased for the true average treatment effects within each partisanship condition. To prove that this estimator is unbiased, we need to show that its expected value is equal to the true average treatment effect (ATE) within the partisanship condition.

   Expectation Operator:

$$E(SATE_{\text{Republican}}) = E\left(\frac{1}{N_{\text{Republican}}} \sum_{i=1}^{N_{\text{Republican}}} (Y_{\text{Jake},i} - Y_{\text{DeShawn},i})\right)$$

$$E(SATE_{\text{Democrat}}) = E\left(\frac{1}{N_{\text{Democrat}}} \sum_{i=1}^{N_{\text{Democrat}}} (Y_{\text{Jake},i} - Y_{\text{DeShawn},i})\right)$$

Linearity of Expectation:

$$E\left(SATE_{\text{Republican}}\right) = \frac{1}{N_{\text{Republican}}} \sum_{i=1}^{N_{\text{Republican}}} E\left(\left(Y_{\text{Jake},i} - Y_{\text{DeShawn},i}\right)\right)$$

$$E\left(SATE_{\text{Democrat}}\right) = \frac{1}{N_{\text{Democrat}}} \sum_{i=1}^{N_{\text{Democrat}}} E\left(\left(Y_{\text{Jake},i} - Y_{\text{DeShawn},i}\right)\right)$$

Unbiasedness of Treatment Effect:

$$E\left(SATE_{\text{Republican}}\right) = \frac{1}{N_{\text{Republican}}} \sum_{i=1}^{N_{\text{Republican}}} \tau_{\text{Republican}}$$

$$E\left(SATE_{\text{Democrat}}\right) = \frac{1}{N_{\text{Democrat}}} \sum_{i=1}^{N_{\text{Democrat}}} \tau_{\text{Democrat}}$$

Simplify:

$$E\left(SATE_{\text{Republican}}\right) = \tau_{\text{Republican}} \frac{1}{N_{\text{Republican}}} \sum_{i=1}^{N_{\text{Republican}}} 1$$

$$E\left(SATE_{\text{Democrat}}\right) = \tau_{\text{Democrat}} \frac{1}{N_{\text{Democrat}}} \sum_{i=1}^{N_{\text{Democrat}}} 1$$

Simplify Further:

$$E\left(SATE_{\text{Republican}}\right) = \tau_{\text{Republican}}$$

$$E\left(SATE_{\text{Democrat}}\right) = \tau_{\text{Democrat}}$$

Therefore, we have shown that, under the assumption of complete randomization, the estimators for SATE within each partisanship condition are unbiased for the true average treatment effects within those conditions.

7. The output:

```
No Partisanship Signal:
{'observed_statistic': 0.02064040421530533, 'p_value': 0.202}

Democratic Signal:
{'observed_statistic': -0.0034981179269369944, 'p_value': 0.845}

Republican Signal:
{'observed_statistic': -0.017086906048901174, 'p_value': 0.267}
```

- For the No Partisanship Signal, the p-value is 0.202, suggesting that there's moderate evidence against the null hypothesis, but it doesn't reach conventional significance levels.
- For the Democratic Signal, the p-value is 0.845, indicating that there's no strong evidence against the null hypothesis.
- For the Republican Signal, the p-value is 0.267, suggesting that there's moderate evidence against the null hypothesis, but, again, it doesn't reach conventional significance levels.

These results imply that, based on the observed data, there isn't strong evidence to reject the null hypothesis of a zero causal effect for any of the tested conditions.

8.

```
SATE for Republican Legislators: 0.005524861878453308
SATE for Democratic Legislators: 0.0022371364653244186
```

Both Republican and Democratic legislators show a positive effect when moving from the DeShawn condition to the Jake condition under no partisanship signal. However, these effects are relatively small (around 0.5% and 0.2% increase, respectively). The patterns suggest a mild impact of the treatment within each subgroup, indicating a nuanced influence of the treatment on legislators' responses.

9.

1. Republican legislators had a larger racial gap in response, as the SATE for Republican legislators (0.0055) is numerically larger than the SATE for Democratic legislators (0.0022).

2. In the causal inference framework used in this class, establishing causation requires addressing potential confounding variables and ensuring the validity of the identification strategy. The observed difference in the racial gap between Republican and Democratic legislators may be influenced by various factors other than their political party affiliation. Without accounting for these potential confounders and considering the assumptions underlying causal inference methods, we cannot definitively attribute the observed difference in the racial gap to the political party of legislators. Therefore, within the

context of the methods used, we cannot assert a causal relationship between political parties and the observed racial gap.

10. Block randomization is a randomization technique used in experimental design to ensure balance or control over specific variables or covariates that might influence the outcome. In this method, the experimental units are first divided into blocks based on certain characteristics, and then randomization is applied separately within each block.
    they blocked the following covariates:

    State: Legislators were grouped by the state they represent.
    Legislative Chamber: The authors considered the legislative chamber to which the legislator belonged (e.g., House or Senate).
    Political Party: Legislators were grouped based on their political party affiliation.
    Reelection Status: The authors took into account whether the legislator was up for reelection.
    By blocking on these covariates, the authors aimed to ensure that each treatment group (DeShawn and Jake conditions) had a balanced representation of legislators with similar characteristics within each block. This helps control for potential confounding variables and enhances the internal validity of the experiment.

10.2 Questions

1. Identifying the key variables:

1. **Treatment group**: Electoral districts affected by the Elbe flood

2. **Treatment variable**: Elbe flood

3. **Pre-treatment period:** 1998 - 2002

   Post-treatment period: 2002 - 2009

4. **Outcome variable(s)**:

Short-term: immediate change in SPD PR vote shares after the flood

Long-term: sustained difference in SPD PR vote shares after the flood

2. Difference-in-means (DiM) compares average outcomes between groups at a single point in time. Difference-in-difference (DiD) is a more robust approach to this type of problem and data, and is more plausible because of these reasons:

a. **Time-varying confounders:**

   DiD considers time-varying confounders by comparing changes in outcomes over time within both the treatment and control groups. This also allows the researchers to examine the dynamic policy effects in the short-term and long-term periods.

   Empirical evidence:

   The study clearly states that it is studying the voting effects of the 2002 Elbe flood on the SPD party over the 1998-2009 election periods. Policy effects in short-term and long-term periods can be found in Figures 2 and 3, respectively.

b. **Spatial variables**

Spatial variables can also be accounted for in DiD.

Empirical evidence:

Figure 5 accounts for what electoral districts are actually treated by the flood effect.

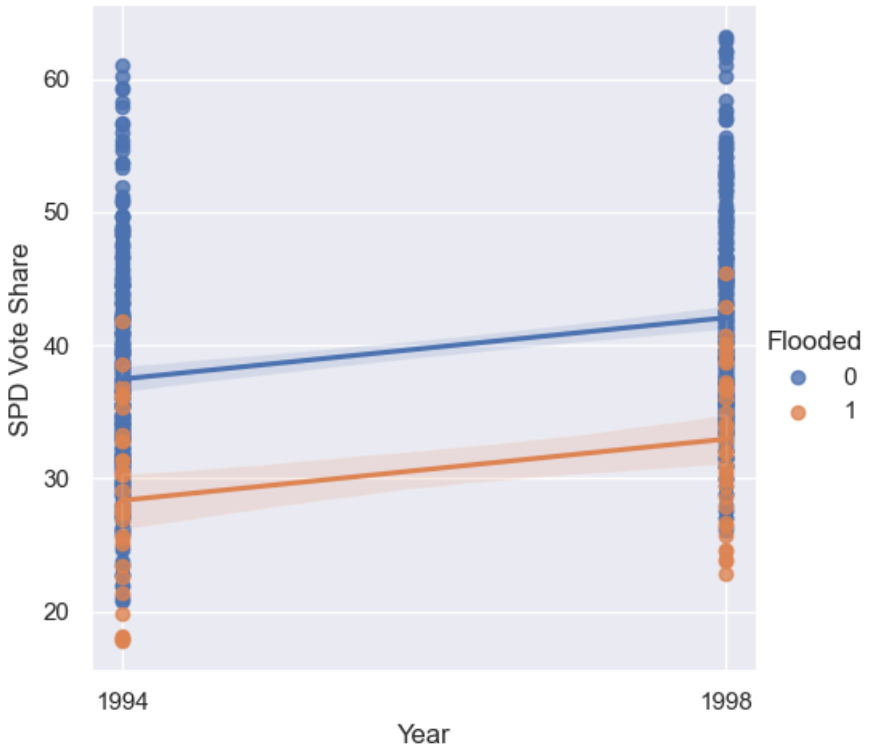
c. **Fixed characteristic confounders:**

DiD controls for other confounders by differencing out time-invariant characteristics.

Empirical Evidence:

The other confounders that are accounted for can be found in Table 1. Also, in Figure 4, policy effects from the Iraqi war are controlled for. Figure 7 explains whether there was actual persuasion or mobilization from the flood effects.

3. The treatment group is the electoral districts by flood and the control group is the electoral districts unaffected by the flood. They checked the plausibility by conducting a placebo test in the pre-treatment period, 1994 and 1998 electoral district elections, and comparing the trends of the treatment and control groups. Plot produced by code in the Section2.ipynb file:



4. Main quantities of interest:

    a. Placebo electoral results:

    Formal Definition: The difference in would-be-treated electoral district vote shares for the SPD over the 1994 and 1998 elections.

    Interpretation: These results verify the parallel trends assumption required for DiD by comparing two different election periods right before the treatment.

    b. Short-term electoral rewards:

    Formal Definition: The immediate increase in vote shares for the incumbent SPD party in directly affected areas in response to the policy response to the 2002 Elbe flooding.

Interpretation: This represents the short-term impact of the beneficial policies on electoral outcomes, capturing the immediate gratitude or approval from voters in the flooded regions.

c. Long-term electoral rewards:

Formal Definition: The persistence of increased vote shares for the incumbent SPD party in directly affected areas in subsequent elections (e.g., 2005 election).

Interpretation: This reflects the longevity of voter gratitude or continued support for the incumbent party due to the policy response to the flooding, extending beyond the short-term impact.

d. Very long-term electoral rewards:

Formal Definition: The prolonged effect on SPD vote shares from the flood response is found by comparing the 1998 election, before treatment, and the 2009 election, which is long after the treatment.

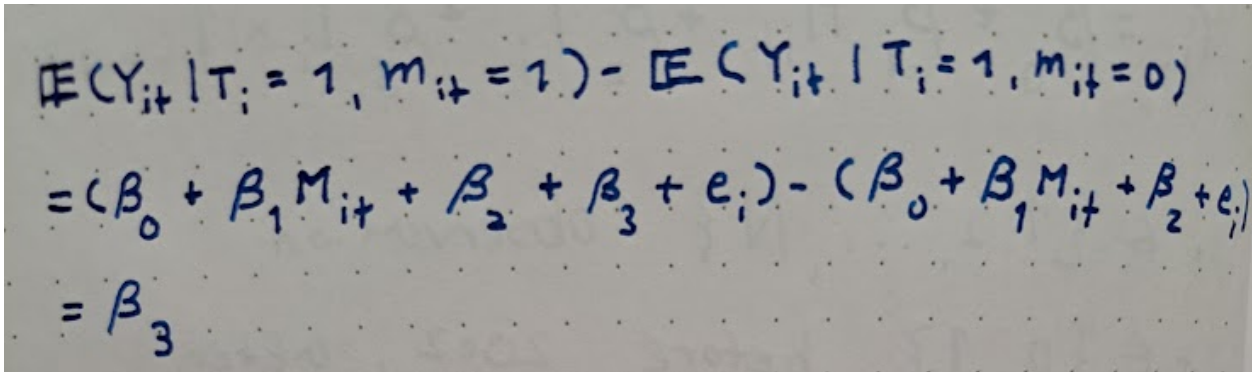
Interpretation: Like the long-term electoral rewards, this quantity showcases the extended voter gratitude, if it still exists, from the SPD response to the Elbe flood.

Means can be found in the Section2.ipynb file under the “part_2” folder.

5. Difference-in-difference estimate:

$$\mathbb{E}[Y_{i,1}(1) - Y_{i,1}(0) \mid T_i = 1]$$

This can be rewritten in the given terms:



Given the abridged definition of Y_{it}.

10.3 Questions

All answers are in the Jupyter Python notebook that can be found in the code under the “part_3” folder.