midterm2

November 12, 2022

1 Midterm 2

This miderm is based on the real-world hiring test that the Analyst Institute used to screen their data analysts. Analyst Institute is a consulting firm that works with progressive and Democratic groups in conducting randomized field experiments.

For this midterm, you will be analyzing an experiment the Analyst Institute conducted.

In 2017, one of their partners was interested in determining the effectiveness of an SMS (text message) persuasion program on likely voters. The campaign hoped their text messages would both persuade voters to support candidate Jane Smith over her opponents, and mobilize voters to cast a ballot in their state's 2017 general election. Voters were randomly assigned to receive one of two interventions:

- No contact (Control group)
- Message reminding them to vote in the upcoming election, and a message highlighting Smith's record on environmental issues (Treatment group)

After the SMS program was implemented, Analyst Institute conducted a live phone survey asking voters which candidate they planned to support in the upcoming election. After the election, they consulted publicly-available state voter files to measure whether the targeted voters actually voted on Election Day. Note that while the state voter files (and thus, our measure of turnout) include everyone in the treatment universe, we only surveyed a subset of the universe. **For the purposes of this midterm, do not worry about implementing survey weights.**

The four main research questions were:

1. Was the experiment properly implemented (balance check between treatment and control on the pre-treatment covariates)?
2. Did the message program increase voters' likelihood to vote?
3. Did the message program persuade voters to support Jane Smith?
4. Did the message program increase how much voters care about protecting the environment on a scale from 1-10?

Before turning in your work, please re-run all of the code and download as a PDF. Review this PDF to make sure everything is printing properly. Once you are finished, please upload to Canvas.

This midterm is open book, notes, prior labs, and Internet. You can use any resource except for another live human being.

You have the entire 110 minutes of class time to complete this exam. Good luck!

To get started, you will need the file `data` for this midterm. Below we load this file and inspect it.

[3]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("data.csv")
```

[4]:
```
data.head()
```

[4]:

	ai_id	race	age	gender	marital_status	sms_treat	turnout2017	support_smith	environment_thermometer
0	f57737fc9c3dabde23271345714dfb6a	black	73	female	separated	Control	1	1	7
1	974f0de9c5860103c481c7dc5d0b9ea2	black	26	female	married	Control	0	0	5
2	544ca018f0e8c73061d7e6eb9739ca94	white	36	male	married	Treatment	0	0	10
3	91f51410c89a3fb11c2bf953dd9686d5	white	39	male	separated	Treatment	0	1	9
4	f8f08aa57b4b724468aa278fe8637377	black	47	male	separated	Control	0	1	7

1.1 Part 1: Balance Check

Does the treatment group look similar to the control group? Conduct a balance check on the pre-treatment covariates. A complete answer will provide a table and a few sentences answering the question (interpretation).

1.2 Part 2: Did the message program increase voters' likelihood to vote?

Answer the question. A complete answer will provide a numerical answer, a visualization, and a few sentences answering the question (interpretation). You do not need to worry about statistical significance when answering this question.

1.3 Part 3: Did the message program persuade voters to support Jane Smith?

Answer the question. A complete answer will provide a numerical answer, a visualization, and a few sentences answering the question (interpretation). You do not need to worry about statistical significance when answering this question.

1.4 Part 4: Did the message program increase how much voters care about protecting the environment on a scale from 1-10?

Answer the question. A complete answer will provide a numerical answer, a visualization, and a few sentences answering the question (interpretation). You do not need to worry about statistical significance when answering this question.