MKT680: Home Assignment 1
Due at 4:00pm on February 13, 2024

You received the data after running an experiment:

1. "transaction_data.csv" – includes the transaction data where the profit was already computed and aggregated on the customer-date level
2. "cust_data.csv" – includes customer id, home state, and treatment group.

The experiment contained two treatment conditions:

1. "Mail". These customers were sent a catalog with the new spring fashion collection
2. "No Mail". These customers were not sent anything and were left as a control group

The experiment started on early morning 2nd March 2019 (the moment when the catalogs were shipped). The randomization was done on the Customer ID level and contained 12,000 representative customers. Each catalog costs $20 to produce and mail to the customer. The total active customer base is 312,000 customers. You have a strong belief that the campaign could not be worse than nothing so you decided to use one-sided tests. Historically your firm tests results at 5% significance level.

1. Before the analysis, you need to preprocess the data.
    a. Calculate the pre-treatment variable "Profit 60 days before the treatment date" for each customer in the experiment
    b. Calculate the outcome "Profit 60 days after the treatment date" for each customer in the experiment
    c. For each of these two variables report the mean and standard deviation.

For the next analysis, your table should look like:

Customer ID	Treatment Group	State	Profit 60 days after the treatment date	Profit 60 days before the treatment date

Check yourself: the table should contain 12,000 rows, the average for "Profit 60 days before the treatment date" is around 77.33, and the average for "Profit 60 days after the treatment date" is around 119.35. Didn't you miss some customers? Is it possible that some customers in the experiment did not purchase anything during the observational period?

2. **Before and After**. Assume you tried to estimate the effect of the campaign without the experiment by using only the treatment group (before and after)
    a. What would be your treatment effect of mailing a catalog?
    b. Assuming the results are statistically significant would you launch a full-scale marketing campaign based on these results?
    c. What is the expected gain of your marketing campaign if you launch it on the remaining population?
    Note: Gain here is part of the total profit attributed to the marketing campaign – total profit with sending a catalog minus total profit without sending a catalog
3. **Randomization check**. Using the methods discussed in class, check the internal validity of the experiment:

   a. Check the categorical variable (State). Which test you would use? Report p-value
   b. Check the "Profit 60 days before the treatment date". Which test you would use? Report p-value
      Note: typically, you can't assume the sign of the difference for the randomization check, so it is better to use a two-sided test.
   c. What is your conclusion, can you proceed to experiment evaluation?
4. **Average Treatment Effect**. Using the methods discussed in class compute the results of the experiment for "Profit 60 days after the treatment date.
   a. Estimate the Average Treatment Effect of sending a catalog
   b. Determine if the results are statistically significant at 5%. Report p-value and standard error
   c. Would you launch a full-scale marketing campaign based on these results?
   d. What is the expected gain of your marketing campaign if you launch it on the remaining population?
   e. How would you compare the results with (2)? Explain.
5. **Difference-in-difference**. You want to use the pre-treatment information to improve the estimates. You decided to use a difference-in-difference estimator:
   a. Estimate the treatment effect by using the diff-in-diff method
   b. Determine if the results are statistically significant at 5%. Report p-value and standard error
   c. Would you launch a full-scale marketing campaign based on these results?
   d. What is the expected gain of your marketing campaign if you launch it on the remaining population?
   e. How would you compare the results with (4)? Explain.
6. **Basic Targeting.** Can you come up with a solution to earn money?:
   a. You have 3 different states in the data. Compute ATE for each state separately.
   b. Would you launch a full-scale marketing campaign based on these results in any of the states?
   c. What is the expected gain of your marketing campaign if you launch it on the remaining population but you don't have to launch it in all states?
      Hint: your random sample is representative of the whole population. Can you tell how many people would be from GA in the remaining 300,000 customers?
7. Compare results in (2), (4), (5) and (6). Write a short recommendation to your manager about the marketing campaign and justify it. (*Short: less than 5 sentences*)