Homework 1: Data Analytics Recap

The purpose of this homework is to test your knowledge on the data analytics methods. For this assignment, upload your report and your code. Feel free to use Excel, R, or Python. In case of the latter, please upload your ipynb file.

Exercise 1: Game Fun: Customer Acquisition through Digital Advertising

Disclaimer: “No similarity to actual persons or companies is intended or should be inferred.”

Game Fun is one of the world's top developers of casual mobile games and spends millions of dollars every year on digital advertising. Of particular interest to Game Fun is their efforts on improving their customer acquisition. It is a common belief in the mobile gaming industry that the more paid traffic you have, the better your organic traffic will be. This is known as K-Factor (see appendix for an explanation). However, measuring the causal effect of online digital advertising has proven to be an extremely challenging task. Towards this end, Game Fun decided to run an A/B experiment.

Display Banners Experiment

Game Fun ran an online display banners advertising campaign with the primary objective of increasing its sales on gaming subscription packages. To attract new users, the display ad advertised their most popular game and offered the user a promotion of $25 signing-up bonus. The credits would appear in the customer’s game account and could be used to purchase any further in-app features. Based on historical data, a new customer subscription brings a revenue of $37.5 on average. This results in a net inflow of $12.5 after the $25 credit for the users acquired through this promotion.

The A/B experiment worked as follows. Before the start of the digital ad campaign, Game Fun chose two different websites (“content publishers”) to run the experiment on. The content publishers have randomly assigned their web users to test and control groups. As users browsed on the two websites, the advertising servers checked whether a given user should be show a Game-Fun ad. If the user qualified for a Game-Fun ad, then the ad server checked whether the user was assigned to the test or the control group. If the user belonged to the test set, a

Game-Fun ad was displayed to the user. Otherwise, a completely irrelevant ad was displayed to the user.

However, Game-Fun had to pay the content publishers for these irrelevant ads, as well. This raised a tension between the management and the data scientist teams. The management team had two concerns. First, paying for other companies ads is directly decreasing their marketing budget. Second, they didn't like the fact that some users who saw an irrelevant ad might have signed up for the Game Fun game in the first place, had they been shown their gaming ad (indirect effect – opportunity cost missed). For these two reasons, the management team asked their data scientist team to carefully decide on the best proportion of users to assign to the test and control groups, while at the same time maintaining a statistically valid comparison.

The data scientist team ran a statistical power analysis of the experiment, and decided to allocate 70% of users to test group and the rest 30% to the control group (see appendix for an explanation on power analysis).

Analysis of the Experiment

After the completion of the experiment, the results came in and you (the data analytics team) is being asked to analyze them. The data are provided in the file "GameFun.xls". Each row in the data belongs to an individual customer. The first column is the anonymized customer id. The second column, "test", indicates whether the user was part of the treatment group (test =1) or the control group (test=0). There are three demographic variables, "gender" (male =1, female =0), "income" (this is measured in thousands), and "gamer" (gamer = 1, if user is a gaming enthusiast; 0, otherwise). The website that a customer visited is in the variable "site". The variable "impressions" contains the number of advertising impressions that a customer received. If a customer is in the test group, then all of this customer's impressions are for the Game-Fun's ad; if a customer is in the control group, then all of this customer's impressions are for the irrelevant ad. Last, the column "purchase" is the dependent variable, and it indicates if the customer purchased anything within 30 days after her/his conversation to the game (30 days is the expected customer lifetime duration for a mobile gamer). If a customer purchased, then purchased =1; 0, otherwise.

Questions

1. Before evaluating the effect of an experiment, it is important to make sure that the experiment was executed correctly. Check whether the test and control groups are probabilistically equivalent on their observables?

a. More specific, compare the averages of the income, gender and gamer variables in the test and control groups. You should also report the % difference in the averages. Compute its statistical significance. **[2 pts]**
b. Briefly comment on what these metrics tell you about probabilistic equivalence for this experiment. **[2 pts]**
c. If you had run this type of analysis BEFORE executing an experiment and found a large difference between test and control groups, what you should do? **[5 pts]**
d. (Open/Ended Question) If you had millions of consumers, your "classic" statistical significance tests would not work (this is because the number of samples is used to compute those classic statistical tests). Do some research online and propose what significance test would you do in case you had "big data"? **[5 pts]**

2. Evaluate the average purchase rates in the test and control for the following groups. For each comparison, report the average purchase rate for the test, average purchase rate for the control and the absolute difference (not the % difference) between the test and control.
    a. Comparison 1: All customers **[2 pts]**
    b. Comparison 2: Male vs Female customers **[2 pts]**
    c. Comparison 3: Gamers vs Non-Gamers Customers **[2 pts]**
    d. Comparison 4: Female Gamers vs Male Gamers **[2 pts]**
3. Assess the expected revenue in the test vs. control for the following comparisons:
    a. Comparison 1: All customers **[4 pts]**
    b. Comparison 4: Female Gamers vs Male Gamers **[4 pts]**
4. Based on your previous answers, provide a brief recommendation to your management team summarizing the expected financial outcome for Game-Fun.
    a. Should Game-Fun run this promotion again in the future? If no, explain why. If yes, should Game-Fun offer it to all customers or a targeted segment. **[10 pts]**

Appendix

K-Factor Explained

Assume you have a market (e.g., a US state or a country) where you have almost zero daily organic downloads. You agree to buy users though targeted advertisement. Let's say that you buy 10,000 users. If after a specific period of time, you suddenly have 12,000 users, you know that 2000 of them came as an indirect result of your paid user acquisition campaign. So, your K-Factor is 1.2. A K-Factor higher than 1 is considered viral, and the higher the K factor, the better.

Statistical Power

In order to determine the fraction of users that need to be assigned to the control group, a data scientists needs to consider three factors: 1) baseline conversion rate, 2) campaign reach and 3) expected minimum advertising lift. See here for more info and a calculator: https://www.optimizely.com/sample-size-calculator/

Over time, students have frequently asked me about the appropriate sample size for control and treatment groups in A/B experiments. When an equal number of units is assigned to both groups, it is considered a "balanced" experiment. However, in some situations, it may be appropriate to allocate fewer units to the control group and more to the treatment group, such as when running a marketing campaign where your managers expect the campaign to have a significant success. Conversely, there are times when it's more advantageous to have more units in the control group and fewer in the treatment group, as in mobile gaming, where you experiment with a major change in the design of an established game. This is known as an "unbalanced" design and has significant trade-offs. Generally speaking, an unbalanced experiment may result in larger confidence intervals for estimation. For further reading, check out this great article: link.

Exercise 2: Non-Compliance in Randomized Experiments

Randomized Experiments

Randomized experiments are the gold standard in making inferences about causal questions. As we learned in the class, sometimes randomized experiments are not feasible in practice due to practical and/or ethical reasons. However, even when they are feasible, complications might arise, which can severely bias our analysis. In what follows we will study one of the real-world limitations in experimental research: non-compliance. Non-compliance happens when subjects take different treatment from their assigned one. In other words, subjects decide to behave differently due to self-selection.

One sided Non-compliance

Oregon Health Insurance Experiment (OHIE) is one of the most well-known randomized controlled trials (RCTs) in health economics, designed to study the effects of Medicaid

expansion. However, the experiment had substantial non-compliance**,** making it a textbook case for instrumental variable (IV) estimation. A group of economists examined the impact of extending access to Medicaid among a low-income, uninsured adult population on health outcomes. Due to limited funds, and because the state (correctly) anticipated that the demand for the program among eligible individuals would far exceed the available slots, Oregon held a lottery to randomly assign Medicaid eligibility to some applicants. However, not all individuals who won the lottery enrolled in Medicaid, while none of those who lost the lottery were eligible to enroll. This is an example of what we call “one-sided non-compliance”.*

**In reality, the Oregon Health Insurance Experiment featured “two-sided non-compliance”, but for the sake of keeping things simple we will consider a variation of this setting with only one-sided non-compliance, as just described above.*

Problem Setting

A policymaker is interested in understanding the impact of Medicaid on health outcomes and financial security. To do so, they consulted four data scientists, each of whom suggested a different method of analysis. Your task is to use the "ohie_data" dataset and implement their recommendations to determine which is the most reliable approach.*

**These are not the data used (nor the results found) in the real study, but a fictitious dataset that we have created for the purpose of this exercise.*

Data Description

The data contains three variables: instrument (equals one if the individual won the Medicaid lottery, zero otherwise), treatment (equals one if the individual enrolled in Medicaid, zero otherwise) and outcome (equals one if the individual experienced a negative health outcome, zero otherwise).

Questions

1. The first data scientist suggested comparing the negative health outcome rates of individuals who won the Medicaid lottery to the negative health outcome rates of individuals who lost the lottery.

    a. What percentage of individuals who won the Medicaid lottery experienced a negative health outcome? **[3 pts]**

b. What percentage of individuals who lost the Medicaid lottery experienced a negative health outcome? **[3 pts]**
c. What is the difference in negative health outcome rates? Under what assumptions is this difference a valid estimate of the causal impact of enrolling in Medicaid on health outcomes? **[4 pts]**

2. The second data scientist suggested comparing negative health outcome rates between those who actually enrolled in Medicaid and those who did not.

    a. What percentage of individuals who enrolled in Medicaid experienced a negative health outcome? **[3 pts]**
    b. What percentage of individuals who did not enroll in Medicaid experienced a negative health outcome? **[3 pts]**
    c. What is the difference in negative health outcome rates? Under what assumptions is this difference a valid estimate of the causal impact of enrolling in Medicaid on health outcomes? Are these assumptions likely to hold in reality? Why (provide concrete examples)? **[4 pts]**

3. The third data scientist advised that one should consider only those who won the Medicaid lottery and compare those who enrolled in Medicaid to those who did not.

    a. What percentage of individuals who enrolled in Medicaid experienced a negative health outcome? **[3 pts]**
    b. What percentage of individuals who won the lottery but did not enroll in Medicaid experienced a negative health outcome? **[3 pts]**
    c. What is the difference in negative health outcome rates? Under what assumptions is this difference a valid estimate of the causal impact of enrolling in Medicaid on health outcomes? **[4 pts]**

4. The fourth data scientist suggested the following Wald estimator to estimate the effect of Medicaid enrollment on health outcomes:

$$\frac{\%\ of\ lottery\ winners\ who\ \ experienced\ a\ negative\ outcome\ \ -\ \%\ of\ lottery\ losers\ who\ experienced\ a\ negative\ outcome}{\%\ of\ lottery\ winners\ who\ \ enrolled\ in\ Medicaid}$$

    a. Compute the above Wald estimate for the given dataset. **[2 pts]**

b. Under what assumptions is this estimate a valid measure of the causal impact of Medicaid on health outcomes? Be specific (i.e. discuss the assumptions in *this* setting). **[4 pts]**

c. On standard errors:
   i. What is the standard error for the intent-to-treat estimate recommended by the first data scientist? What is the standard error for the Wald estimate recommended by the fourth data scientist? **[5 pts]**
   ii. Which one is larger and why? **[4 pts]**
   iii. Why might these standard errors be biased? **[2 pts]**

d. The fourth data scientist also does not take our word for granted and wants to check that there is indeed only one-sided noncompliance. Using the dataset provided, how can they conclude that there are no cases of two-sided noncompliance? In other words, what would they have had to observe in our dataset to detect cases of two-sided noncompliance? **[3 pts]**

Exercise 3: Reading Time

Select any paper from the Canvas folder "Extra" → "Seminal AI Papers" and read it carefully. Then, write a short reflection (up to one page) addressing the following questions. You can structure your response in paragraphs or use bullet points. Your reflection should demonstrate deep engagement with the paper, rather than just summarizing it.

1. Key Insights: What are the most important ideas from the paper? Why do they matter?
2. Impact & Evolution: How do you think this paper has influenced modern AI research? Are there later works that built on it?
3. Limitations & Critiques: What assumptions does the paper make? Are there any limitations that stand out to you?
4. Connections: How does this paper relate to what you've learned in other classes or research projects?
5. Application & Experimentation: If you were to design an experiment inspired by this paper, what would it be?
6. Unanswered Questions: What questions did the paper leave you with?

Important: To encourage original thinking, your response should focus on aspects that require critical analysis rather than just summarization. This exercise is worth 10 points.