An Evaluation of the Causal Effect of a Promotion Campaign on Software Usage

Febriany Lete
Maysen Pagan

April 2024

1 Introduction

In an era marked by the digital transformation of businesses, startups are increasingly reliant on strategic outreach efforts to both attract new customers and improve loyalty among existing ones. One such startup, specializing in software solutions, is looking to increase consumption among existing customers. This startup aims to boost usage among its existing customers by offering discounts. To gauge the impact of these discounts on customer usage, the company examines historical data from 2000 customers. The key variables in the dataset include a binary treatment variable indicating whether a customer received a discount and a response variable representing the revenue generated from software purchases in the year following the discount. Additionally, there are several pre-treatment covariates such as the number of PCs used by the customer, the company's employee count, the amount spent on IT purchases, the company's size (measured by total yearly revenue), and binary indicating factors which include whether the customer has global offices, is a large consumer, is a Small Medium Corporation, or has commerical business.

2 Exploratory Data Analysis

Before performing any causal analysis, we first explore the data for any missing values or outliers. There are no missing values in our data set, however, from a histogram of our response variable, we can see there exists one value that is negative. Since it is unknown if the revenue from this customer was input incorrectly and there is only one observation where the revenue is less than 0, we will remove this customer from our data set.

Figure 1. Histogram of Revenue Response Variable

Next, we look at a correlation matrix of our quantitative variables. From Table 1, we can see a strong linear association of 0.95 between the number of PCs a customer has and the number of employees a customer has. Similarly, there is a large correlation of 0.88 between the size of the customer, given by their yearly total revenue, and the dollars spent on IT purchases by the customer. As a result, we choose to omit the dollars spend on IT purchases and the number of PCs the customer has from further analysis.

1

Table 1: Correlation Matrix of Reponse and Quantitative Predictors

	IT Spend	Employee Count	PC Count	Size	Revenue
IT Spend	1.00	0.02	0.02	0.88	0.74
Employee Count	0.02	1.00	0.95	0.02	0.23
PC Count	0.02	0.95	1.00	0.03	0.24
Size	0.88	0.02	0.03	1.00	0.86
Revenue	0.74	0.23	0.24	0.86	1.00

3 Methods and Analysis

3.1 Assumptions

We then consider the three main identification assumptions to identify our average treatment effect:
Positivity, Consistency, and Conditional Exchangeability.

3.1.1 Positivity

For the Positivity assumption, we observed the covariate balance
plots for each of our covariates, comparing the distribution of the
values between the treatment and control. All plots show the pat-
tern of an equal distribution of covariate values between those cus-
tomers who received the discount and those who did not. For ex-
ample, the covariate balance plot on the right for the employee
count variable displays a skewed right distribution with the count
ranging from 0 to about 500 employees both for customers who re-
ceived a discount and customers who did not receive a discount.
This demonstrates that for given values of each of the covariates,
the probability of receiving the treatment and control are both non-
zero. Therefore,

$$0 < p(A = a|X) < 1, \ a \in \{0, 1\}$$

3.1.2 Consistency

We will also perform our causal analysis under the Consistency
assumption which states that a customer's potential outcome under
the observed treatment assignment is equal to the observed outcome
for that customer. Therefore,

$$\text{If } A_i = a, \text{then } Y_i = Y_i^{(a)}$$

3.1.3 Conditional Exchangeability

For the Conditional Exchangeability assumption, we thought of possible variables that could be a
confounder of the treatment-outcome relation. We believe that the size of the customer could be this
confounder. The startup company might tend to give discounts to larger companies either to keep

Employee Count Under Treatment

Employee Count Under Control
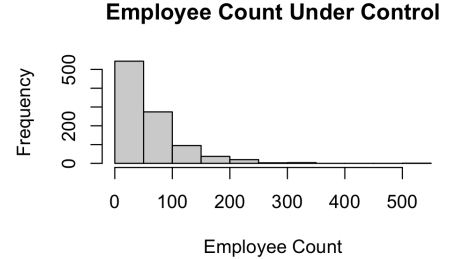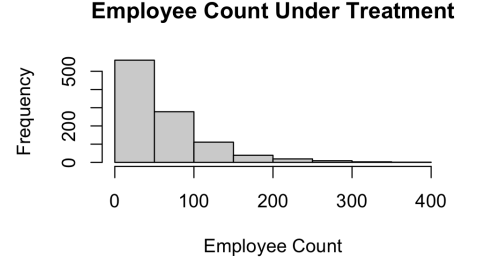
their loyalty or prevent them from purchasing from their competitors. In this sense, the size of the customer directly affects the treatment assignment. Additionally, the size of the customer may also directly affect the outcome because, for example, larger customers will most likely purchase more software than smaller customers and therefore bring in more revenue for the company. However, this size of the customer confounder can be observed in our data set through variables such as the number of employees the customer has or the binary variable that indicates if the customer has global offices. Larger businesses tend to have more employees as well as global offices. Since this confounder can be observed in our data set, we continue with the assumption of Conditional Exchangeability which states that

$$\{Y^{(0)}, Y^{(1)}\} \perp\!\!\!\perp A | X$$

3.2 Estimation

Under the above assumptions, which are known as the three main identification assumptions, our parameter of interest, the average treatment effect (ATE), can be identified and then estimated using three methods.

We first use the simplest estimator for the ATE which is the Outcome Regression (OR) Estimator. We also choose to estimate $\hat{\mu}$ for the OR function using linear regression with ordinary least squares estimator with $(1, A, X)$ as predictors. We then use non-parametric bootstrap to generate 1000 ATEs and estimated the standard error and confidence intervals for the OR estimator from these bootstrap estimates.

$$\hat{\theta}_{ATE}^{or} = \mathbb{E}_n[\hat{\mu}(1, X) - \hat{\mu}(0, X)]$$

The Inverse Probability Weighting (IPW) approach, which models the treatment assignment mechanism instead of the outcome mechanism, is not invariant to location transformation of the outcome and also is not stable leading to high variance. Therefore, we also calculate the ATE using the Hajek estimator which is invariant to location transformation of the outcome and is more stable (although this means it also has increased bias). Non-parametric bootstrap was used to estimate standard error and obtain confidence intervals.

$$\hat{\theta}_{ATE}^{hajek} = \mathbb{E}_n\left[\frac{\frac{A}{\hat{\pi}(X)}}{\mathbb{E}_n\left[\frac{A}{\hat{\pi}(X)}\right]}Y - \frac{\frac{1-A}{1-\hat{\pi}(X)}}{\mathbb{E}_n\left[\frac{1-A}{1-\hat{\pi}(X)}\right]}Y\right]$$

The OR function in the OR Estimator represents the outcome mechanism while the propensity score in the IPW and Hajek Estimators represents the treatment assignment mechanism. We also estimate the ATE using the Doubly Robust (DR) Estimator which combines both information and is unbiased if either the OR function or the propensity score is correctly specified. Similar to the previous estimators, non-parametric bootstrap was used to estimate standard error and obtain confidence intervals.

$$\hat{\theta}_{ATE}^{hajek} = \mathbb{E}_n\left[\frac{A\{Y - \mu(1, X; \hat{\beta})\}}{\pi(X; \hat{\alpha})} + \mu(1, X; \hat{\beta}) - \frac{(1-A)\{Y - \mu(0, X; \hat{\beta})\}}{1 - \pi(X; \hat{\alpha})} - \mu(0, X; \hat{\beta})\right]$$

The propensity score was estimated using logistic regression with an order 1 polynomial. For each of the covariates, the estimated propensity scores were used to calculate $\frac{1}{n}\sum_{i=1}^{n}\{\frac{A_i}{\hat{\pi}(x_i)} - 1\}X_{i_p}$ for $p = 1,2,...,6$ for 6 predictors. Bootstrap was used to calculate 1000 of these values for each covariate

and 95% confidence intervals were produced. The results are shown in Table 2 below. From this table, we can see that the value 0 falls in every confidence interval which suggests that our estimator of the propensity score satisfies the weighted balancing property for each of the covariates.
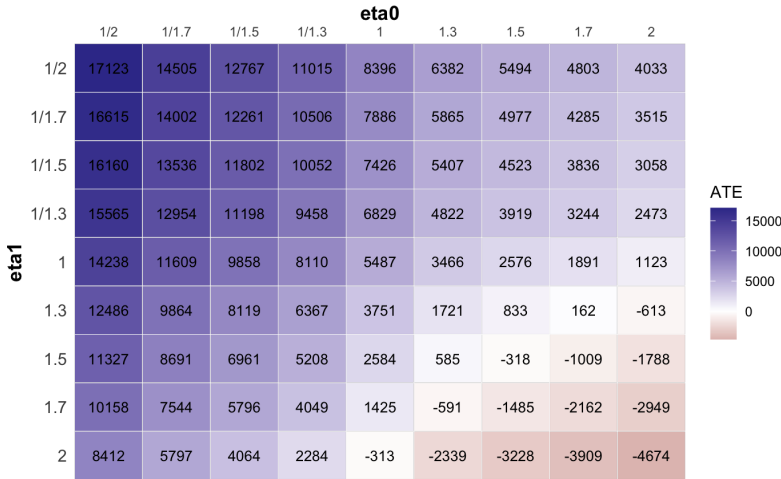
Table 2: Confidence Intervals for Covariate Balance Check

Covariate	Lower Bound	Upper Bound
Global Flag	-0.04	0.05
Major Flag	-0.03	0.06
SMC Flag	-0.06	0.07
Commercial Flag	-0.07	0.09
Employee Count	-6.77	7.52
Size	-8984.85	10340.57

The figures below display the ATE estimates as well as the bootstrap confidence intervals using the three methods above. From these figures, we can see that all of the 95% bootstrap confidence intervals do not contain zero suggesting that there is a positive causal effect between giving a customer a discount and the revenue received from that customer. We focus on the DR estimator because only one of the OR function or propensity score needs to be correctly specified for the estimator to be unbiased. We observe that the ATE is about 5,500 which means that on average, customers who received the discount would give the company \$5,500 more in revenue than those who did not receive the discount.
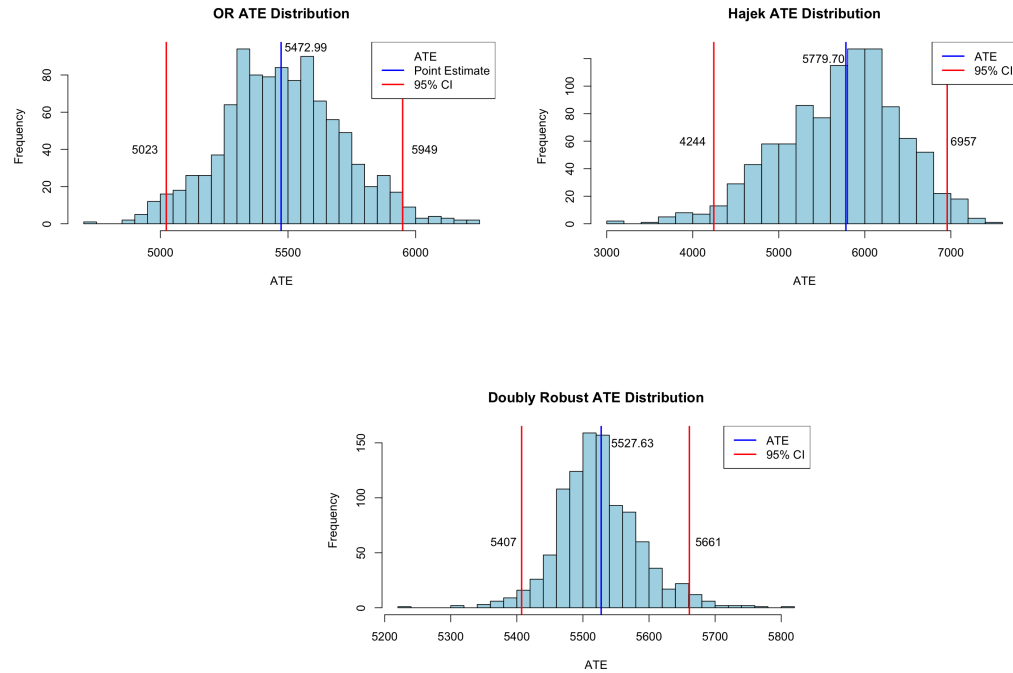
3.3   Sensitivity Analysis

We then performed sensitivity analysis to observe the extent to which the Conditional Exchangeability assumption might be violated since this assumption can not be tested. We chose our sensitivity parameters to range from $\frac{1}{2}$ to 2 and the figure below is a heatmap of the calculated effects with certain combinations of these parameters. We can see from the bottom right corner of the heatmap that the signs of our ATEs are sensitive only for larger sensitivity parameters. This suggests that when customers who receive a discount tend to bring in more revenue (which means the sensitivity parameters are larger than 1), then our results are sensitive. On the other hand, for customers who receive a discount and tend to bring in less revenue, our results are not sensitive and we can conclude that there is a positive causal effect.

		eta0							
eta1	1/2	1/1.7	1/1.5	1/1.3	1	1.3	1.5	1.7	2
1/2	17123	14505	12767	11015	8396	6382	5494	4803	4033
1/1.7	16615	14002	12261	10506	7886	5865	4977	4285	3515
1/1.5	16160	13536	11802	10052	7426	5407	4523	3836	3058
1/1.3	15565	12954	11198	9458	6829	4822	3919	3244	2473
1	14238	11609	9858	8110	5487	3466	2576	1891	1123
1.3	12486	9864	8119	6367	3751	1721	833	162	-613
1.5	11327	8691	6961	5208	2584	585	-318	-1009	-1788
1.7	10158	7544	5796	4049	1425	-591	-1485	-2162	-2949
2	8412	5797	4064	2284	-313	-2339	-3228	-3909	-4674

ATE
- 15000
- 10000
- 5000
- 0

4   Conclusion

Our conducted causal analysis revealed a significant increase of approximately \$5,500 in revenue among customers who received the discount compared to those who did not. This indicates a positive average treatment effect of the discount on revenue generation. Through sensitivity analysis, we conclude that our estimated causal effect is primarily sensitive to only higher levels of unmeasured confounding. Despite this sensitivity, our analysis provides confidence in the existence of a causal effect, although the accuracy of the estimated ATE may be subject to some degree of inaccuracy.

From this analysis, we learned the importance of visualizing and exploring the data prior to applying any causal inference frameworks. The data may contain missing or unusual values that may affect the analysis without our knowledge. Additionally, we learned how to apply the causal inference methods learned to real data. Exploring and understanding the data allowed us to determine which assumptions are satisfied and then apply a method that utilizes those assumptions to identify and estimate the average treatment effect.