

TITLE*
SUBTITLE

Adrian Wong, Yingying Zhou, Xinyi Xu, Yang Wu

26 February 2021

Abstract

ABSTRACT

Contents

| | | |
|----------|---------------------------------|-----------|
| 1 | Introduction | 2 |
| 2 | Data | 2 |
| 2.1 | Intervention | 2 |
| 2.2 | Data Gathering Method | 2 |
| 2.3 | Descriptive Analysis | 2 |
| 3 | Discussion | 7 |
| 3.1 | Overview | 7 |
| 3.2 | Findings | 7 |
| 3.3 | Future Directions | 8 |
| 4 | Appendix | 9 |
| 4.1 | Appendix A | 9 |
| 4.2 | Appendix B | 9 |
| | References | 10 |

*Code and data are available at: <https://github.com/yanggl224/groupproject-.git>

Table 1: First 6 rows Raw data

| type | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 | Q13 |
|---------|-----|---------|----------------|-----------|----|-----|-----|------|-------|-----|-----------|-----------|-------|
| Control | M5W | Toronto | Family Style | Franchise | 11 | Yes | No | 1-10 | 20.11 | No | No change | No change | 44140 |
| Control | L7C | Peel | Fine Dining | No | 10 | Yes | No | 1-10 | 23.31 | No | No change | No change | 42217 |
| Control | L7C | Peel | Family Style | No | 2 | Yes | Yes | 1-10 | 16.74 | No | No change | Decrease | 37507 |
| Control | L6A | York | Fast Casual | No | 2 | Yes | Yes | 1-10 | 19.21 | No | No change | No change | 41194 |
| Control | L6H | Halton | Premium Casual | No | 1 | Yes | No | 1-10 | 15.22 | No | No change | No change | 56615 |
| Control | L4Z | Peel | Fast Food | No | 3 | Yes | No | 1-10 | 15.60 | No | No change | No change | 51303 |

1 Introduction

2 Data

2.1 Intervention

2.2 Data Gathering Method

2.3 Descriptive Analysis

After discussing data gathering method, we sampled data in R (R Core Team 2020). We totally have **3274** observations, and 14 of following features according to the questionnaires.

- **type** : Categorical identifier [“Treated” or “Control”] for each observation
- **Q1** : First three digits of the postcode
- **Q2** : Categorical identifier for distinguishing the type of restaurants
- **Q3** : Region name in GTA
- **Q4** : Describe whether the restaurant is a franchise (“Franchise” or “No”)
- **Q5** : The length of the operation years for each restaurant
- **Q6** : Describe whether the restaurant offer takeout service (“Yes” or “No”)
- **Q7** : Describe whether the restaurant offer delivery service (“Yes” or “No”)
- **Q8** : Number of employees in the restaurant (category type)
- **Q9** : Average employee hourly rate (CAD)
- **Q10** : Describe whether the restaurant has been a site of a potential COVID case (“Yes” or “No”)
- **Q11** : Describe the restaurant’s fixed costs change situation
- **Q12** : Describe the restaurant’s flexible costs change situation
- **Q13** : The restaurant’s past month revenue (CAD)

The first six rows of raw data is shown in the Table1. (Table 1)

2.3.1 EDA

Taking a deep look at all the features from survey questionnaire, we learned some demographic features about the restaurants in GTA:

- From figure1(Figure 1) and figure2(Figure 2), we noticed that more restaurants are located in Toronto (around 500) and Peel (around 400). The number of restaurants in Hilton is similar to the number in Durham. Meanwhile, Casual dining takes the lead in the restaurant type in GTA, with around 26%. Then it comes to Family style restaurant, accounting for 20%. Fast food restaurant, fine dining restaurant and Premium casual restaurant almost equally make up 10%. There is no big difference between treated group and control group in terms of restaurant number and type distributions.

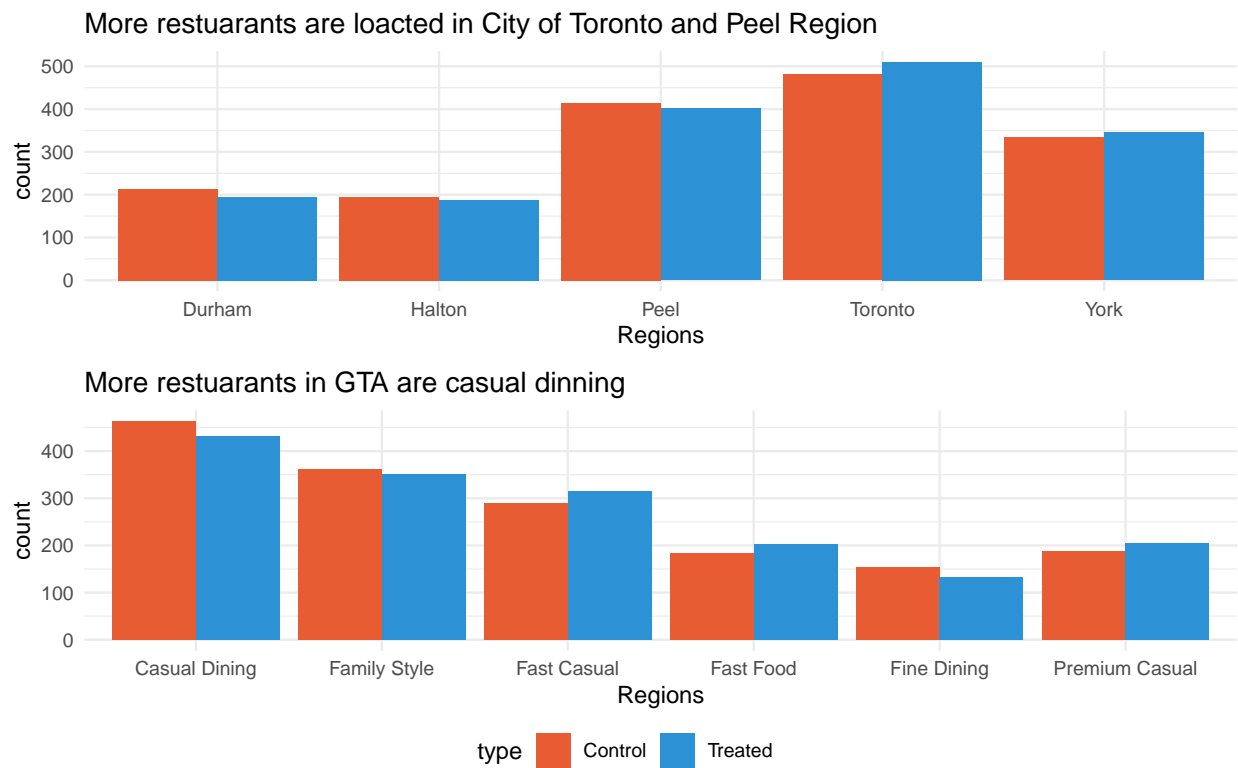


Figure 1: Restaurant numbers and types

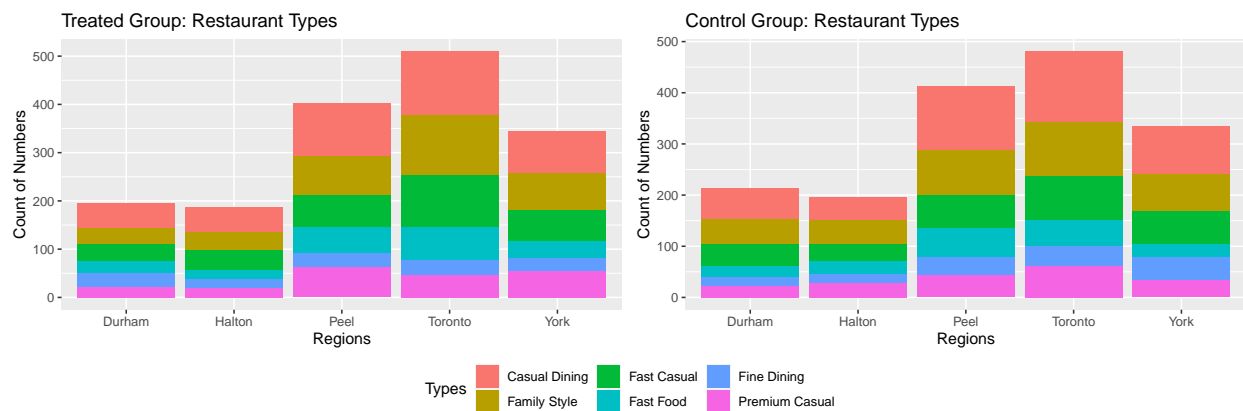


Figure 2: Restaurant types

- In terms of employees salary, the average hourly rate before and after intervention is both around 17 CAD. In contrast with two boxplots, we can see there is a slight increase in the treated group. The reason behind might because the employee take more risks to go for work, accordingly they will receive higher salary. (Figure 3)

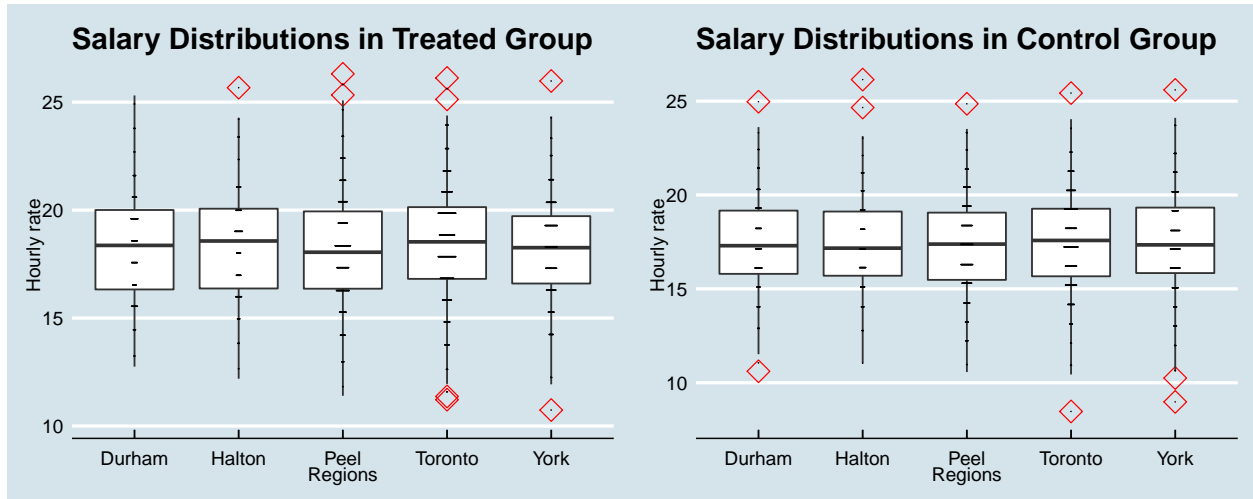
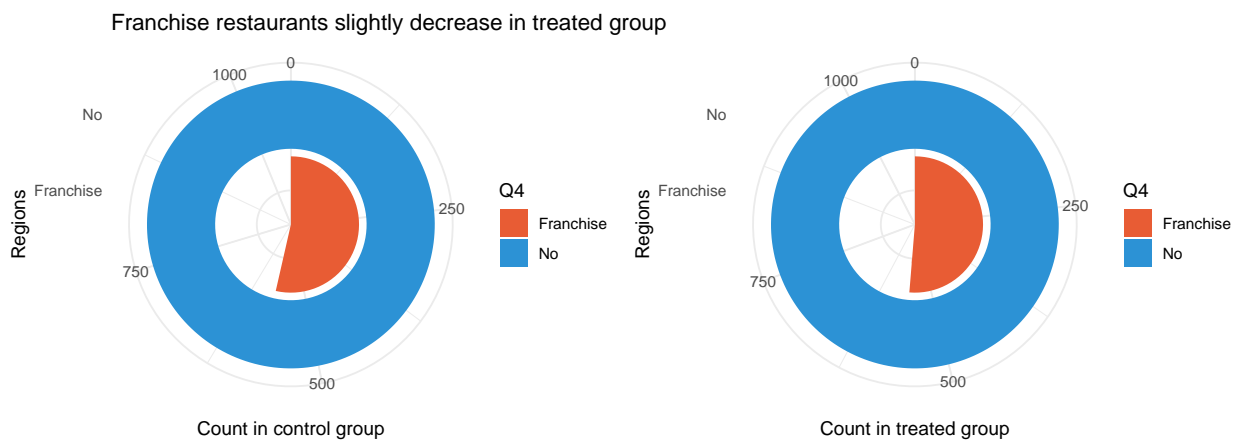


Figure 3: Employee salary distribution

- The attributes of the restaurant determine its management mode, so whether the restaurant is a franchise is quite important factor. From the pie charts, we found the portion of franchise rate in treated group is slightly lower than control group. We assume in treated group, the branch franchise restaurants should follow the rules by the head office. Considering the potential cost of COVID issues, chain restaurants will face greater risks, which is why they are less likely to be in the treated group. (Figure ??)



* The polar chart illustrates the employee numbers distribution, as can be seen in figure below. (Figure 4) Because of COVID rule, no restaurant is allowed to open for large group dine in. So in the control group, there is 0 restaurant which has more than 30 employees. Most of restaurant has 10 to 20 employees.

Almost 0 restaurant has more than 30 employees in control group

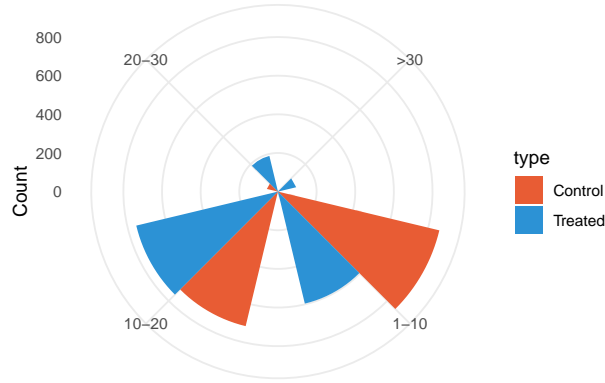


Figure 4: Employee numbers distribution

Table 2: T Test on the Restaurant's revenue

| mean_of_Treated | mean_of_Control | p.value | conf.low | conf.high | method | alternative |
|-----------------|-----------------|---------|----------|-----------|-------------------------|-------------|
| 63143.44 | 49358.97 | 0 | 11638.18 | 15930.76 | Welch Two Sample t-test | two.sided |

2.3.2 T-Test

The T Test is used to compare the sample mean of our Treated group and Control group. The goal is to determine whether the intervention has an effective effect on the treated group. Our hypothesis is the intervention will have positive impact towards the restaurant's revenue. (Kim 2015) The T test results is represented in the Table2(Table 2). The package **Broom**(Robinson, Hayes, and Couch 2021) is used to clean the t test results and convert it into the dataframe. The p value we get is $< 2.2e-16$, as the p value would indicate a significant result, meaning that the actual p value is even smaller than $2.2e-16$ (a typical threshold is 0.05, anything smaller counts as statistically significant).(Kim 2015) So we can interpret hypothesis not rejected which means the intervention has a significant effect on treated group.

2.3.3 Correlation matrix

Correlation matrix shows internal relationships between x variables and y variable. (Figure 5) Intensity is indicated by the color(from red to blue). No significant coefficiency is barred with symbol "x". More detailed analysis will be conducted in finding part.

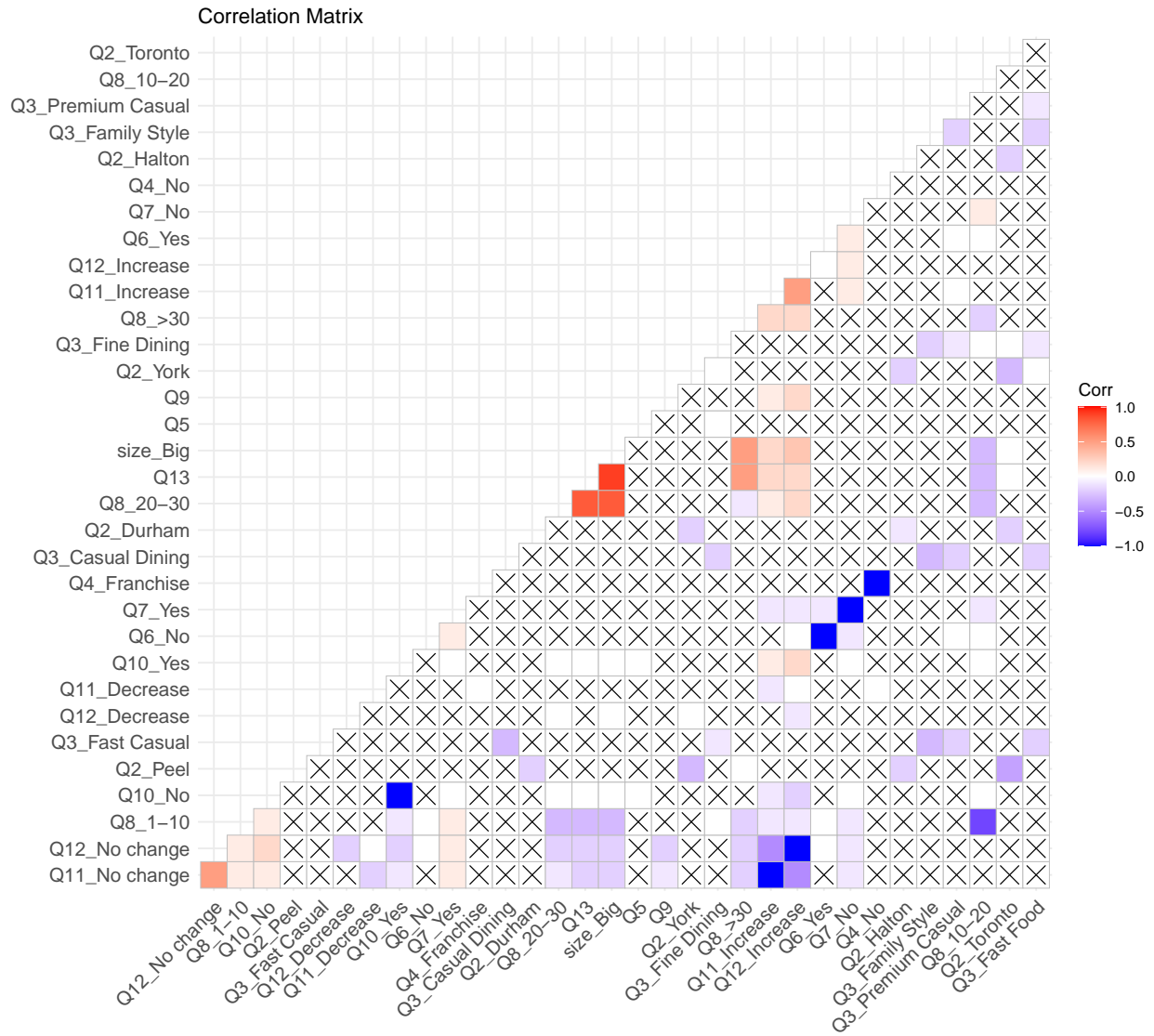


Figure 5: Correlation matrix

3 Discussion

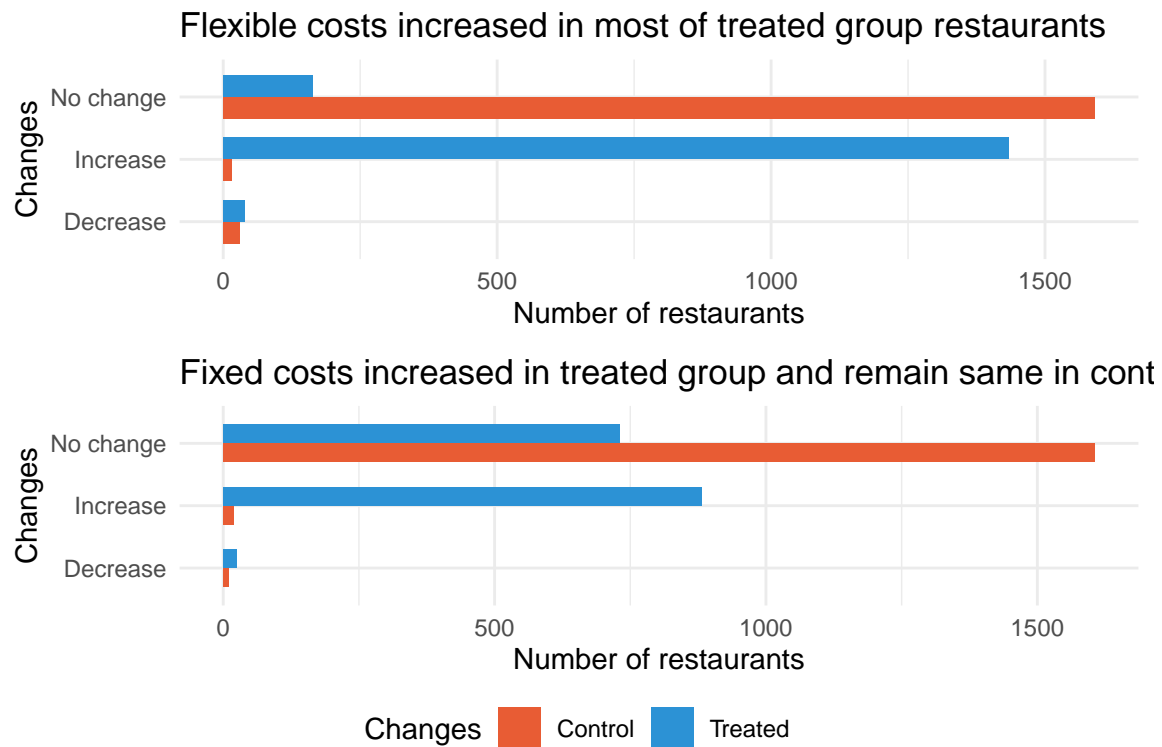
3.1 Overview

3.2 Findings

3.2.1 Finding ONE

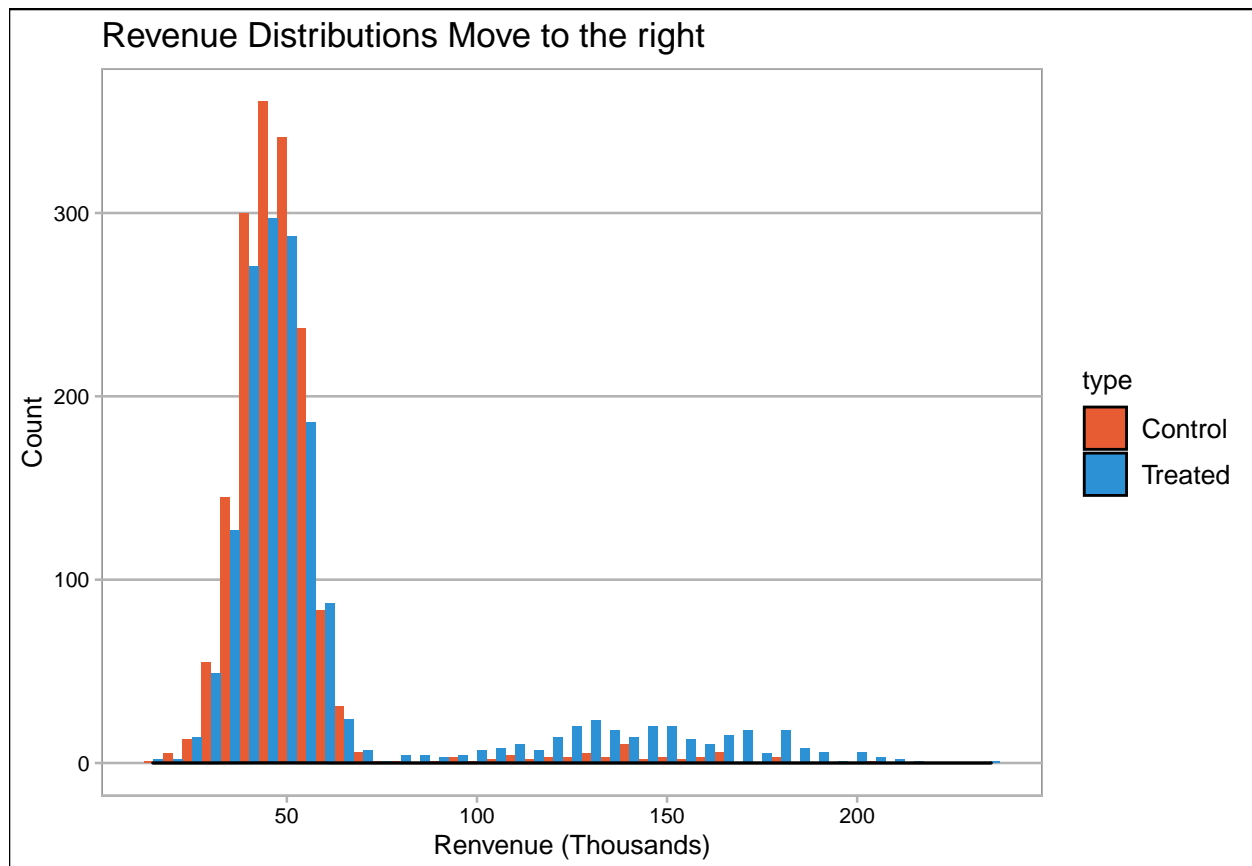
3.2.2 Finding TWO

Invention effect on Flex and fixed cost



###

Finding THREE Invention effect on Revenue distributions



Limitation

3.3 Future Directions

Table 3: Detailed information for stratification

| Region | Number of Restuarants | Proportion(%) | Sample Selected |
|---------|-----------------------|---------------|-----------------|
| Toronto | 7500 | 29.58 | 48430 |
| Durham | 3260 | 12.86 | 21051 |
| York | 5553 | 21.90 | 35858 |
| Peel | 6235 | 24.59 | 40262 |
| Halton | 2803 | 11.06 | 18100 |
| Total | 25351 | 100.00 | 1637 |

Table 4: Estimated Cost

| Components | Cost per unit | Total cost for each component |
|---------------|---------------|-------------------------------|
| Printing Cost | 0.05 | 738.95 |
| Envelope Cost | 0.15 | 4433.70 |
| Stamp Cost | 0.55 | 16256.90 |

4 Appendix

4.1 Appendix A

4.2 Appendix B

References

- Kim, Tae Kyun. 2015. “T Test as a Parametric Statistic.” *Korean Journal of Anesthesiology* 68 (6): 540.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robinson, David, Alex Hayes, and Simon Couch. 2021. *Broom: Convert Statistical Objects into Tidy Tibbles*. <https://CRAN.R-project.org/package=broom>.