

How Starbucks Reward Loyalty Program steering customer's spending - Propensity Score Matching approach

Fanzhen Meng Student No.1002812824
<https://github.com/Fanzhen-Meng/304FinalProject.git>

12/22/2020

Abstract

The propensity-score matching (PSM) technique is usually used to estimate the difference in outcomes between the treatment and control group within a particular program. This paper studies the effect of Starbucks membership on customers' average spending where membership is seen as a treatment. The treatment is assigned based on propensity score which is estimated by giving observed customer characteristics: age, gender, and income. The results showing that the effect of treatment on average spending is significant, and this may provide some insights into Starbucks' marketing strategy.

Keywords: Starbucks membership, observational study, propensity score matching(PSM), causal inference, logistic regression, multiple linear regression(MLR).

Introduction

Starbucks Corp. has rapidly grown into the world's dominant coffee shop chain in the past five decades by roasting, marketing, and selling coffee and ever-expanding merchandise of other food, beverages, and branded products. While Starbucks dominates the U.S. market, it faces increasing competition in international markets, including U.K. Costa Coffee, a subsidiary of Coca-Cola Co.; and China-based Luckin Coffee Inc(Farley, 2020). Under tough market competition, the Starbucks Reward Loyalty Program plays an important role in generating customer loyalty, increasing revenue, collecting data, and creating personalized marketing efforts(Leah, 2020).

According to Starbucks' official data, the Loyalty Program has climbed to 16 million active members as of March 2019(Leah, 2020). The program had attributes 40% of Starbucks total sales and has helped same-store sales rise by around 7%(Brown, 2019). As a member of the Loyalty Program, people can order and pay through the app ahead of arriving at a nearby store, as well as earn stars (rewards) that can be used to get free drinks, food, and merchandise. In order to help the corporations achieve the highest level of customer engagement, it is important to understand how the membership identity and other customer characteristics could contribute to a customer's spending in Starbucks.

When simulating the data, three general customer characteristics "age", "gender", and "income" will be considered as independent variables. In the following study, I will be focusing on investigating the association between customer's membership identity, as well as the three characteristics and their average spending in Starbucks.

Methodology

Data

This study will be an observational study that requires a general process of simulating data. The data set includes a sample size of 10000 participants, and each participant was randomly given four values of independent variables: Age,

Table 1:

id	age	gender	income	membership	average_spend	propensity score
9504	61.89865	Female	below 40k	0	45.42400	0.0022011
7184	61.88365	Female	below 40k	1	54.43570	0.0022035
19	56.32888	Female	below 40k	0	30.66416	0.0033177
8888	56.32728	Female	below 40k	1	49.62527	0.0033181
4244	67.67844	Female	40k-75k	0	32.58191	0.0048242
9156	67.67838	Female	40k-75k	1	44.97306	0.0048243
6977	66.85226	Female	40k-75k	1	49.39384	0.0051263
1555	66.84028	Female	40k-75k	0	45.65204	0.0051308
5879	69.62621	Male	below 40k	1	49.57542	0.0058421
7550	69.60737	Male	below 40k	0	35.21220	0.0058502
2658	64.03806	Female	40k-75k	1	46.63538	0.0063034
5306	64.03650	Female	40k-75k	0	46.43288	0.0063041
5811	62.72982	Female	40k-75k	0	27.36688	0.0069386
5077	67.27024	Male	below 40k	1	52.93766	0.0069450
8394	65.46782	Other/decline	above 75k	1	50.99730	0.0072355
3327	62.15452	Female	40k-75k	0	41.93505	0.0072376

Gender, and Income. Age is a continuous variable with a range from 18 to 100. A ‘runif’ function had been used here to ensure the randomness of the simulating process. Gender has three levels: “Female”, “Male”, and “Other/decline” with randomly assigning probabilities of 0.49, 0.49, and 0.02 respectively. Income has three levels: “below 40k”, “40k-75k”, and “above 75k”.

After generating the sample data, we need to set the probability of being treated with membership depending on the variables. According to a market research report done in 2017, the main target population of Starbucks is men and women in the middle to upper class who can afford the high-priced beverages regularly(Dudovskiy, 2017). The target age of Starbucks’ market ranged from 22 to 60, with the teen customers growing steadily(Mellinger, 2019). Moreover, Starbucks customers are roughly distributed as 50% male, 35% female, and others who are unwilling to disclose their gender(Soicher,2018). Therefore, I believe customers who have memberships are likely to be young males with high incomes. Finally, I expect the outcome of having memberships is spending more at Starbucks on average.

Then a logistic regression model has been constructed to explain the person’s probability of being a Starbucks member, the function was applied to the dataset, which allows me to create matches. A match finds the highest similarities between a Starbucks member and a regular customer based on the propensity score, thus, by examining the matched dataset we have, we are able to explain the effect of having Starbucks memberships on average spending(Note: by ‘average spending’, I mean specifically ‘average weekly spending’). The matching function from the ‘arm’ package in R could be used to do that. Finally, the dataset has been reduced to just those that are matched with 2912 observations, i.e. 1456 pairs of treatment and control participants. Table 1 displays part of the matched dataset.

Model

In order to make causal inferences, subjects are randomly selected and randomly assigned to each group. However, in observational studies, a major limitation is that randomization cannot be achieved or applied even when the participants are randomly selected (Arane, 2008). When there is a lack of randomization, causal inferences cannot be made because it is impossible to determine whether the difference in outcome between two groups is solely caused by the treatment itself or along with other covariates(Arane, 2008). In other words, when subjects with certain traits are more likely to receive the treatment than others, it may lead to a false-positive result that there is an effect of the treatment on the outcome of interests. Hence I will use the propensity score matching technique to analyze the following observational study. The estimated propensity score $e(x_i)$, for subject i , ($i = 1, \dots, N$) is the conditional probability of being assigned

to a particular treatment given observed covariates x_i (Arane, 2008):

$$e(X_i) = Pr(Z_i = 1|X_i)$$

where:

$Z_i = 1$ if the i^{th} participant is in treatment group, and 0 if in control group. X_i is the vector of observed covariates for the i^{th} subject.

To take a step further to introduce the propensity-score matching (PSM) technique, it is usually used to estimate the difference in outcomes between beneficiaries and non-beneficiaries within a particular program (Josiah, 2014). Propensity score value is an estimated probability of the participants receiving the treatment giving their observed characteristics. In this study, having the Starbucks membership is seen as the treatment and the outcome we are interested in is the customers' average spending. The observed covariates are age, gender, and income.

By matching participants who did not get the treatment to those who receive the treatment together, one of the major advantages of PSM is that it can control for confounding variables and extraneous variables and thus reduce biases in observational studies (Josiah, 2014). This is based on the assumption that all variables that affect treatment assignment and outcome have been measured (Austin, 2011). Extraneous variables are variables not manipulated but still have effects on the supposed outcome of the study. A confounding variable is another variable that impacts the independent and dependent variable at the same time, while the effects cannot be differentiated because of its relation to both. For example, in a study where you want to investigate if being a male causes liver cancer, drinking would be a confounding variable (Arane, 2008).

Finally, multiple linear regression (MLR) will be used to analyze the 'effect' of being treated with membership on customers' average spend where treatment is a predictor. Note that in PSM, the casual inferences can only be predicted but cannot be affirmed (Josiah, 2014). The MLR model will be:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon$$

where, for $i=n$ observations, y_i is the dependent variable - average weekly spend; $x_{i1}x_{i2}x_{i3}x_{i4}$ are the predictor variables used to predict the outcome - age, gender, income, and membership. β_0 is the constant term when all of the independent variables are equal to zero; $\beta_1\beta_2\beta_3\beta_4$ are the slop coefficients for each predictor; ϵ is the model's error term.

To be brief, PSM includes four steps: 1) estimating the probability of receiving the treatment, ie. the propensity score, for each participant in the sample data; 2) matching treatment with control participants in order to construct a comparison group; 3) assessing the quality of the resulted matched samples; 4) estimating the treatment effect and then interpreting the results (Josiah, 2014).

Results

Figure 1: Back to back histogram using Hmisc

Propensity score before matching

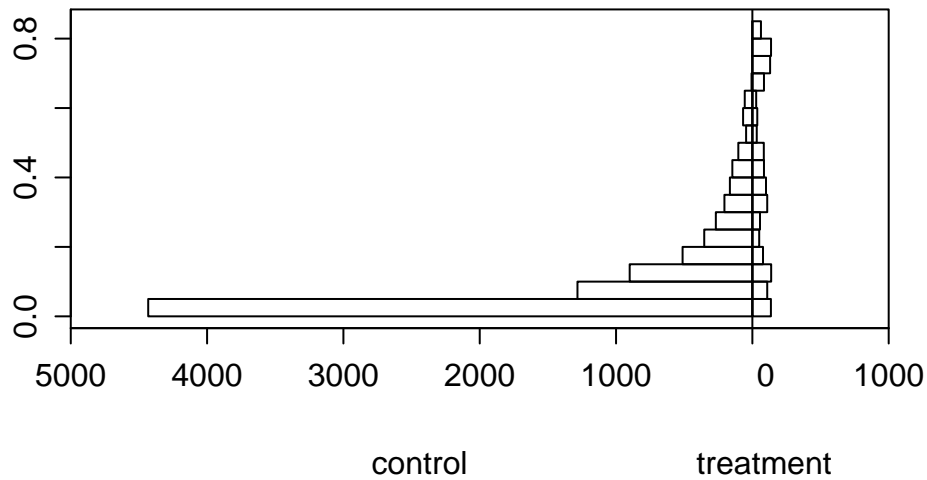
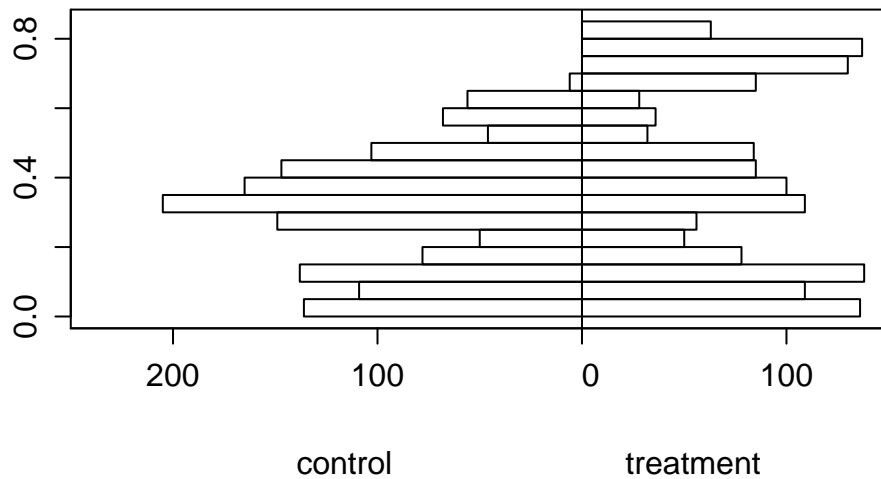


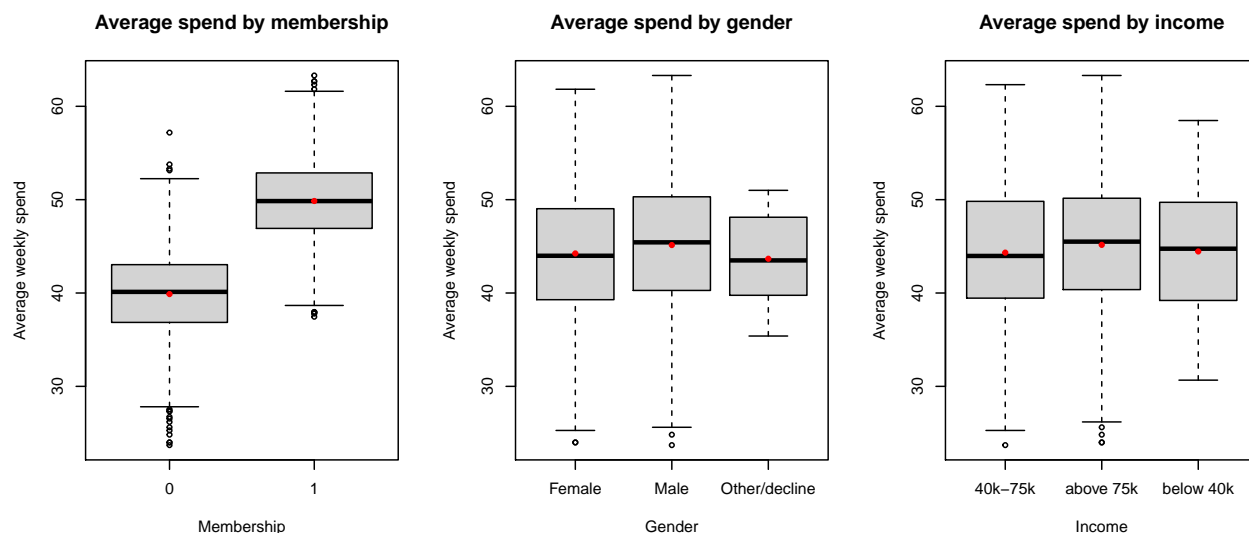
Figure 2: Back to back histogram after match

Propensity score after matching



By comparing figure 1 (before matching) to figure 2 (after matching), we can see that there is a remarkable improvement of the propensity scores after matching. The match suggests that the two groups are much more similar based on the propensity scores. Therefore, selection bias has been reduced substantially.

Figure 3: Boxplots



The boxplots show participant's average spend for the treatment group and control group; for participants classified by gender; and participant classified by income.

From the first plot, we can see that participants who were treated with membership spend approximately 50 dollars weekly on average, which is clearly higher than participants in the control group. Moreover, many outliers can be found outside the whiskers of the box plot. To be specific, most of the outliers in the control group are located at the lower bound, while the outliers in the treatment group are concentrated at the upper bound. The distribution of average spending among people classified by gender(plot 2) and income(plot 3) seems similar in variance and mean(these red dots). In particular, males have a higher average spending on average, but not much. People with income above 75k seem to have a slightly higher average spending.

Table 2: Model Summary

Observations	2912
Dependent variable	average_spend
Type	OLS linear regression

F(6,2905)	589.55
R ²	0.55
Adj. R ²	0.55

	Est.	S.E.	t val.	p
(Intercept)	39.79	0.30	131.36	0.00
age	0.01	0.01	0.77	0.44
genderMale	-0.04	0.19	-0.23	0.82
genderOther/decline	-0.38	1.32	-0.29	0.77
incomeabove 75k	-0.07	0.20	-0.35	0.73
incomebelow 40k	-0.16	0.32	-0.50	0.62
membership1	9.99	0.17	58.71	0.00

Standard errors: OLS

Discussion

Model summary

According to Table 2 summary, the multiple regression model is:

$$y_{\text{spend}} = 39.788 + 0.005(\text{age}) - 0.043(\text{male}) - 0.384(\text{genderOther}) - 0.069(\text{incomeAbove75k}) - 0.161(\text{incomeBelow40k}) + 9.989(\text{membership})$$

The regression coefficient measures the mean change in the dependent variable given a one unit change in the predictor variable. A one-unit increase in age is associated with a 0.005 unit increase in average spend holding gender, income, and membership constant. Participants who were treated with membership have higher average spending, by approximately 9.989 units, holding age, gender, and income constant. For example, we can estimate the weekly average spend in Starbucks of a 40-year-old male, with income below 40k who has a Starbucks membership as follows:

$$y_{\text{spend}} = 39.788 + 0.005(40) - 0.043(1) - 0.161(1) + 9.989(1) = 49.773$$

There is strong statistical evidence(p-value<0.0001) shows that the effect of predictor variable membership on average spending is significant. Whereas other independent variables do not reach statistical significance with p-values 0.84,0.81,0.75,0.62,0.13. This can also be observed in the side-by-side boxplot(Figure 3) where gender and income levels all look similar in variance and mean. Moreover, we can observe that most of the outliers in the control group are located at the lower bound(low average spend) while the outliers in the treatment group are mostly at the upper bound(high average spend). This might also be the evidence that proves the positive effect of treatment on average spending.

Relating model with the world

Since a member of the Loyalty Program can earn rewards and use them to get for free drinks, food, and merchandise, this is a great way for customers to save money at Starbucks(Leah, 2020). More than this, Starbucks is continuously revamping its Reward Loyalty Program to offer more promotions and personalized presents to its members(Palnitkar,2019). Hence the result that participants who were treated with Starbucks membership have 10 dollars higher in average weekly spending is not surprising. Essentially, improving publicity and inviting more customers to the Starbucks Reward Loyalty Program may play a key role in generating customer loyalty, promoting average spend, and increasing revenue.

Limitation/Weaknesses

Because the propensity scores are obtained from observational data, there is no randomization, and the matching only controlled for the differences in the observed variables. Therefore, there may be some bias resulting from the unobserved covariates that may affect whether the participants receive the treatment or not(Arane, 2008). In particular, this study only used age, gender, and income as independent variables, while other relevant covariates are not considered, so this may likely cause issues. Statistically, the data has been used twice. So another weakness in the propensity score-matched sample would be a lack of independence. In addition, the causal inferences can not be made but only be predicted because there is no randomization.

Next Steps

Since covariate balance is a large sample property in propensity score matching(Josiah, 2014), my next step will be combining propensity score matching with regression adjustment(Austin, 2011). By using regression adjustment, the bias that due to residual differences in observed baseline covariates between treatment groups can be reduced(Austin, 2011). The adjustment will ultimately result in a more precise continuous outcome and increase the statistical power of the outcome(Austin, 2011). Finally, a sensitivity analysis would be suitable to determine how robust the results are(Olmos, 2019).

References

- Alexander, Rohan. (2020, November 05). Difference in differences. Retrieved December 23, 2020, from https://www.tellingstorieswithdata.com/06-03-matching_and_differences.html
- Arane, T. Lix, L. (2008, April 22). University of Manitoba. Propensity Score Matching in Observational Studies. Retrieved December 20, 2020, from https://www.umanitoba.ca/faculties/health_sciences/medicine/units/chs/departamental_units/mchp/protocol/media/propensity_score_matching.pdf
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399-424. doi:10.1080/00273171.2011.568786
- Brown, D. (2019, April 16). Starbucks rewards change: New loyalty plan launches today – what you need to know. Retrieved December 10, 2020, from <https://www.usatoday.com/story/money/2019/03/19/starbucks-redesigns-its-loyalty-program-so-you-get-free-stuff-sooner/3174227002/>
- Dudovskiy, John. (2017, April 2). Starbucks Marketing Strategy – Communicating the message of quality via multiple channels. Retrieved December 18, 2020, from <https://research-methodology.net/starbucks-marketing-strategy-3/>
- Farley, A. (2020, Nov 16). How Starbucks Makes Money. Retrieved December 9, 2020, from <https://www.investopedia.com/articles/markets/021316/how-starbucks-makes-money-sbux.asp>
- Josiah Kaplan Oxford University. Oxford. (2014, November 05). Propensity Scores. Retrieved December 20, 2020, from https://www.betterevaluation.org/en/evaluation-options/propensity_scores
- Leah, H. (2020, Feb 9). Starbucks: Winning on rewards, loyalty, and data. Retrieved December 10, 2020, from <https://digital.hbs.edu/platform-digit/submission/starbucks-winning-on-rewards-loyalty-and-data/>
- Mellinger, Barbara B. (2019, Feb 5). Who Is Starbucks' Target Audience? Retrieved December 18, 2020, from <https://smallbusiness.chron.com/starbucks-target-audience-10553.html>
- Olmos, A., & Govindasamy, P. (2015). Propensity Scores: A Practical Introduction Using R. *Journal Of MultiDisciplinary Evaluation*, 11(25), 68-88. Retrieved from https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/431
- Palnitkar, S. (2019, September 16). Starbucks Rewards Case Study - What Makes It Work? Retrieved December 22, 2020, from <https://zinrelo.com/loyalty-rewards-case-study-new-starbucks-rewards-program.html>
- Soicher, P. (2018, November 16). What Type of Starbucks Customer Are You? Retrieved December 19, 2020, from <https://towardsdatascience.com/what-type-of-starbucks-customer-are-you-72f70e45f65>
- Software used in producing the report: Rstudio
- Packages used in producing the report:
- Tidyverse: Easily Install and Load the 'Tidyverse'
- broom: Convert Statistical Objects into Tidy Tibbles
- Hmisc: Harrell Miscellaneous
- KableExtra: Construct Complex Table with 'kable' and Pipe Syntax
- jtools: Analysis and Presentation of Social Scientific Data
- Retrieved December 22, 2020, from https://cran.r-project.org/web/packages/available_packages_by_name.html

Appendix

Data simulating

#Simulating observational data (this code is from Rohan Alexander, <https://www.tellingstorieswithdata.com>)

```
library(tidyverse)
```

```
sample_size <- 10000  
set.seed(304)
```

```
starbucks_data <-  
  tibble(  
    unique_person_id = c(1:sample_size),  
    age = runif(n = sample_size,  
               min = 18,  
               max = 70  
    ),  
    gender = sample(  
      x = c("Female", "Male", "Other/decline"),  
      size = sample_size,  
      replace = TRUE,  
      prob = c(0.49, 0.49, 0.02)  
    ),  
    income = sample(  
      x = c("below 40k", "40k-75k", "above 75k"),  
      size = sample_size,  
      replace = TRUE  
    )  
  )
```

*# Now we need to add some probability of being
treated with membership, which depends on
our variables. Younger, higher-income, male all make it slightly more likely.*

```
starbucks_data <-  
  starbucks_data %>%  
  mutate(age_num = case_when(  
    age < 30 ~ 3,  
    age < 50 ~ 2,  
    age < 70 ~ 1,  
    TRUE ~ 0),  
    gender_num = case_when(  
      gender == "Male" ~ 3,  
      gender == "Female" ~ 2,  
      gender == "Other/decline" ~ 1,  
      TRUE ~ 0),  
    income_num = case_when(  
      income == "above 75k" ~ 3,  
      income == "40k-75k" ~ 2,  
      income == "below 40k" ~ 1,  
      TRUE ~ 0)  
  ) %>%  
  rowwise() %>%  
  mutate(sum_num = sum(age_num, gender_num, income_num),
```



```

      softmax_prob = exp(sum_num)/exp(9),
      membership = sample(
        x = c(0:1),
        size = 1,
        replace = TRUE,
        prob = c(1-softmax_prob, softmax_prob)
      )
    ) %>%
  ungroup()

starbucks_data <-
  starbucks_data %>%
  dplyr::select(-age_num, -income_num, -gender_num,
               -sum_num, -softmax_prob)

# Finally, we need to have some measure of a person's
# average spend. We want those with membership
# to be slightly higher than those without.

starbucks_data <-
  starbucks_data %>%
  mutate(mean_spend = if_else(membership == 1, 50, 40)) %>%
  rowwise() %>%
  mutate(average_spend = rnorm(1, mean_spend, sd = 4.5))
  ) %>%
  ungroup() %>%
  dplyr::select(-mean_spend)

# Fix the class on some
starbucks_data <-
  starbucks_data %>%
  mutate_at(vars(gender, income, membership), ~as.factor(.))
  # Change some to factors
table(starbucks_data$membership)

# Let's take a glimpse at the simulated data
head(starbucks_data)

## Membership is our Treatment.
## Average spending is our outcome of interest.

## Propensity score matching it will be for the membership propensity.

# Now we construct a logistic regression model
# that 'explains' whether a person was treated
# as a function of the variables that we think
# explain it.

propensity_score <- glm(membership ~ age + gender + income,
                        family = binomial,
                        data = starbucks_data)
summary(propensity_score)

```

```

# We will now add our forecast to our dataset.
library(broom)
starbucks_data <-
  augment(propensity_score,
    data = starbucks_data,
    type.predict = "response") %>%
  dplyr::select(-.resid, -.std.resid, -.hat, -.sigma, -.cooksdi)

# Now we use our forecast to create matches.
# For every person who was actually treated (given
# membership) we want the untreated person who
# was considered as similar to them (based on
# propensity score) as possible.

starbucks_data <-
  starbucks_data %>%
  arrange(.fitted, membership)

# Here we're going to use a matching function
# from the arm package. This finds which is the
# closest of the ones that were not treated, to
# each one that was treated.

starbucks_data$treated <-
  if_else(starbucks_data$membership == 0, 0, 1)

starbucks_data$treated <-
  as.integer(starbucks_data$treated)

matches <- arm::matching(z = starbucks_data$treated,
  score = starbucks_data$.fitted)

starbucks_data <- cbind(starbucks_data, matches)

# Now we reduce the dataset to just those that
# are matched. We had 302 treated, so we expect
# a dataset of 704 observations.

starbucks_data_matched <-
  starbucks_data %>%
  filter(match.ind != 0) %>%
  dplyr::select(-match.ind, -pairs, -treated)

head(starbucks_data_matched)

# Examining the 'effect' of being treated on average
# spend in the 'usual' way.

propensity_score_regression <-
  lm(average_spend ~ age+ gender+ income+ membership,
    data = starbucks_data_matched)
summary(propensity_score_regression)

```

Table 1

```
starbucks_data_matched <- starbucks_data_matched[-8]
colnames(starbucks_data_matched)[colnames(starbucks_data_matched) == 'unique_person_id'] <- 'id'
colnames(starbucks_data_matched)[colnames(starbucks_data_matched) == '.fitted'] <- 'propensity score'

library(kableExtra)
short_data <- head(starbucks_data_matched, n=16)

short_data %>%
  kbl(caption = "") %>%
  kable_classic(full_width = F, html_font = "Cambria")
```

Figure 1

```
#(This code is from Olmos, A., & Govindasamy, P. (2015) https://journals.sfu.ca/jmde/index.php/jmde_1/a
library(Hmisc)
histbackback(split(starbucks_data$.fitted, starbucks_data$membership), main= "Propensity score before matching")
```

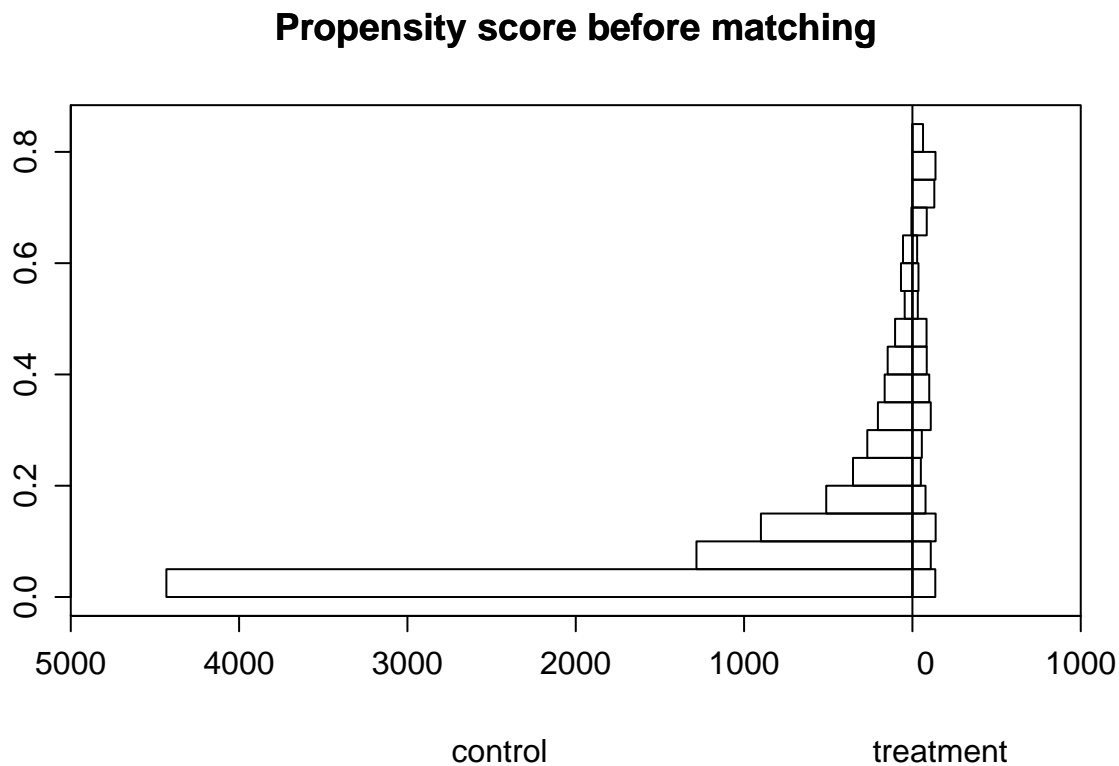


Figure 2

```
library(Hmisc)
histbackback(split(starbucks_data_matched$"propensity score", starbucks_data_matched$membership), main= "Propensity score after matching")
```

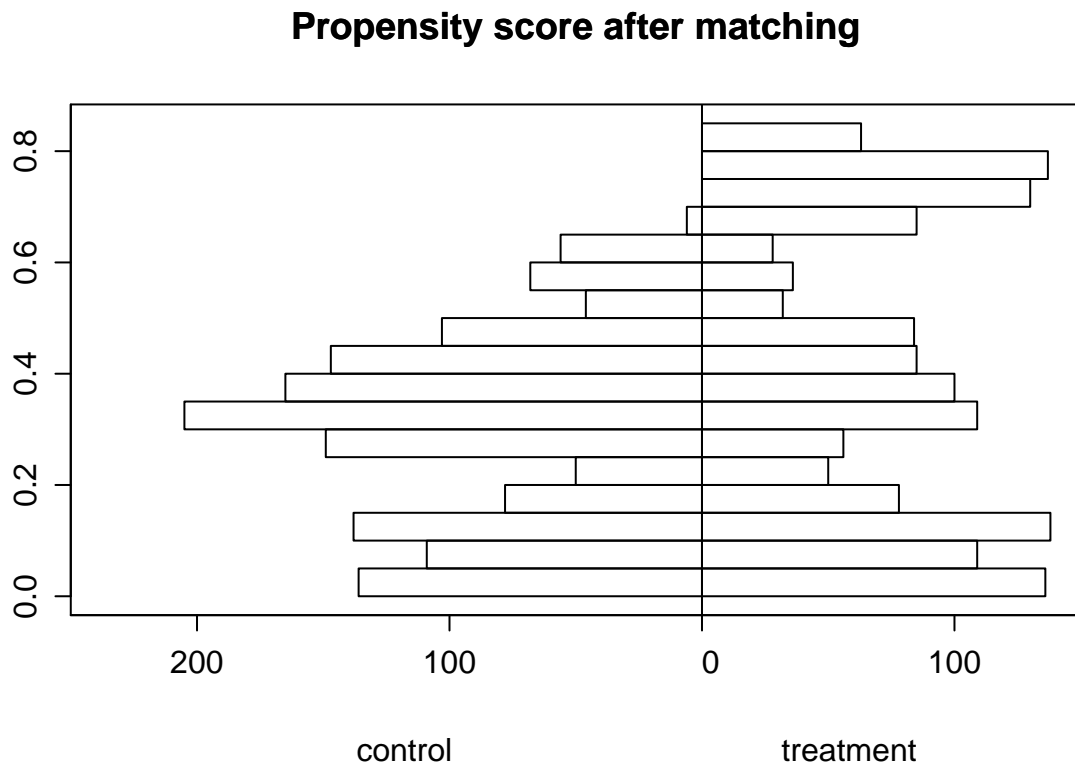


Figure 3

```
par(mfrow=c(1,3))
boxplot(starbucks_data_matched$average_spend~starbucks_data_matched$membership, main="Average spend by membership",
membership_mean <- tapply(starbucks_data_matched$average_spend,starbucks_data_matched$membership,mean)
points(membership_mean,col="red",pch=20)

boxplot(starbucks_data_matched$average_spend~starbucks_data_matched$gender, main="Average spend by gender",
gender_mean <- tapply(starbucks_data_matched$average_spend,starbucks_data_matched$gender,mean)
points(gender_mean,col="red",pch=20)

boxplot(starbucks_data_matched$average_spend~starbucks_data_matched$income, main="Average spend by income",
income_mean <- tapply(starbucks_data_matched$average_spend,starbucks_data_matched$income,mean)
points(income_mean,col="red",pch=20)
```

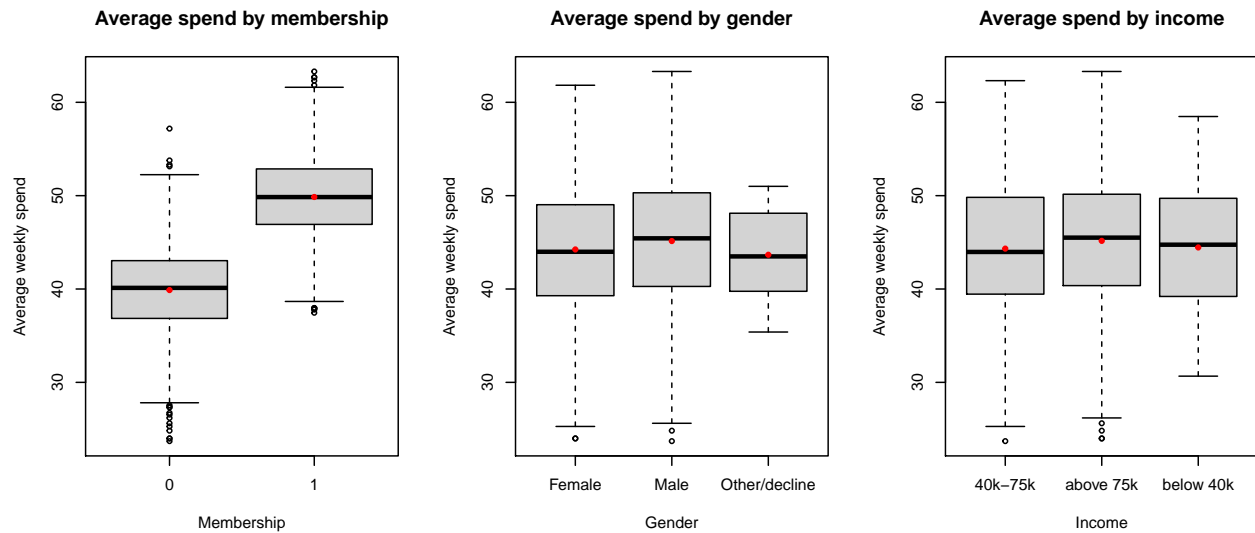


Table 2

```
library(jtools)
summ(propensity_score_regression)
```

Observations	2912
Dependent variable	average_spend
Type	OLS linear regression

F(6,2905)	589.55
R ²	0.55
Adj. R ²	0.55

	Est.	S.E.	t val.	p
(Intercept)	39.79	0.30	131.36	0.00
age	0.01	0.01	0.77	0.44
genderMale	-0.04	0.19	-0.23	0.82
genderOther/decline	-0.38	1.32	-0.29	0.77
incomeabove 75k	-0.07	0.20	-0.35	0.73
incomebelow 40k	-0.16	0.32	-0.50	0.62
membership1	9.99	0.17	58.71	0.00

Standard errors: OLS