The effect of having a "sustainably sourced" badge on customer likelihood of conversion

Maria Stella Vardanega, Deniz Ipek Ozakyol, Lujain Alqassar, Tavishi Pandey, Yipeng (Caroline) Guo

11/29/2021

Introduction

In recent years, there has been an increasing focus on sustainability. The lock downs caused by the COVID-19 pandemic have highlighted the effects of unsustainable supply chains and carbon emissions as was seen from the drastic decrease in air pollution during these lock downs (Rume & Islam, 2020). We wanted to assess whether offering a sustainable product would significantly affect consumer preference in a more sustainable direction. If yes, this may also guide related companies to save energy, reduce emissions and develop green products as they would be catering to the consumer market and subsequently could gain more profits.

Research has shown that when an item is produced sustainably the sales grow 5.6 times faster compared to one that isn't. Furthermore, the market share of products that claim to be sustainably sourced has been growing since 2013 (Whelan & Kronthal-Sacco, 2019). Unilever has also stated that its "sustainable living" brand makes up 70% of its turnover growth. This information portrays that consumers are, in fact, purchasing more sustainable products.

With our experiment we wanted to see whether having a "sustainably sourced" label on the product would significantly impact consumer preference towards that product.

Research Question: Does a sustainably sourced badge on a product cause an increase in consumer preference towards that product

Hypothesis: The presence of a "sustainably sourced" badge on a product increases consumer preference towards that product.

```
survey <- fread('BA830-Group-Project.csv')
```

```
#Removing unnecessary columns and rows
survey <- survey[-c(1:2),]
survey <- survey[, -c(3,4,7,10:13,16,17)]
```

```
#Renaming control columns
names(survey)[9] <- "Control 1"
names(survey)[13] <- "Control 2"
names(survey)[20] <- "Control 3"
```

```
#Replacing all empty strings with NA values
survey[survey==""] <- 0
for(z in 9:20) {
  survey[[z]] <- sapply(survey[[z]], as.numeric)
}
```

```
#Creating two lists that indicate whether the participant was part of the treatment and the outcome
treatment <- c()
outcome <- c()
```

```r
arm <- c()
gender <- c()
age <- c()
major <- c()
for(m in colnames(survey[,9:20])) {
  i <- 1
  while(i <= nrow(survey)) {
    if(m %in% c('Control 1', 'Control 2', 'Control 3')) {
      if(survey[[m]][i] != 0) {
        treatment <- c(treatment, 0)
        outcome <- c(outcome, 0)
        gender <- c(gender, survey[['Q15']][i])
        age <- c(age, survey[['Q16']][i])
        major <- c(major, survey[['Q17']][i])
        arm <- c(arm, 0)
      }
    } else if((m == 'Treatment 1') && (survey[[m]][i] != 0)) {
      arm <- c(arm, 1)
      if(survey[[m]][i] == 1) {
        treatment <- c(treatment, 1)
        outcome <- c(outcome, 1)
        gender <- c(gender, survey[['Q15']][i])
        age <- c(age, survey[['Q16']][i])
        major <- c(major, survey[['Q17']][i])
      } else {
          treatment <- c(treatment, 1)
          outcome <- c(outcome, 0)
          gender <- c(gender, survey[['Q15']][i])
          age <- c(age, survey[['Q16']][i])
          major <- c(major, survey[['Q17']][i])
      }
    } else if((m == 'Treatment 2') && (survey[[m]][i] != 0)) {
        arm <- c(arm, 2)
        if(survey[[m]][i] == 2) {
          treatment <- c(treatment, 1)
          outcome <- c(outcome, 1)
          gender <- c(gender, survey[['Q15']][i])
          age <- c(age, survey[['Q16']][i])
          major <- c(major, survey[['Q17']][i])
        } else {
            treatment <- c(treatment, 1)
            outcome <- c(outcome, 0)
            gender <- c(gender, survey[['Q15']][i])
            age <- c(age, survey[['Q16']][i])
            major <- c(major, survey[['Q17']][i])
        }
    } else if((m == 'Treatment 3') && (survey[[m]][i] != 0)) {
        arm <- c(arm, 3)
        if(survey[[m]][i] == 3) {
          treatment <- c(treatment, 1)
          outcome <- c(outcome, 1)
          gender <- c(gender, survey[['Q15']][i])
          age <- c(age, survey[['Q16']][i])
```

```r
          major <- c(major, survey[['Q17']][i])
        } else {
            treatment <- c(treatment, 1)
            outcome <- c(outcome, 0)
            gender <- c(gender, survey[['Q15']][i])
            age <- c(age, survey[['Q16']][i])
            major <- c(major, survey[['Q17']][i])}
      }
    i <- i + 1
  }
}
```

```r
#Creating the data table for the regression
regression_df <- data.frame(treatment, outcome, arm, gender, age, major)
regression_table <- as.data.table(regression_df)
```

Methodology

Procedure

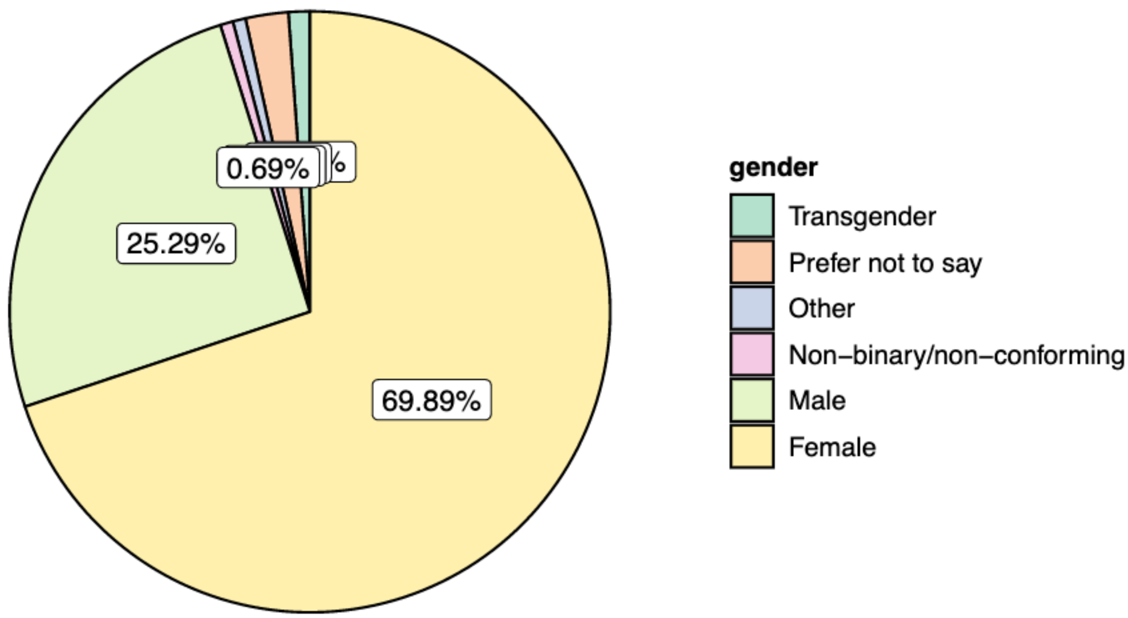
We first decided what products to use in our experiment. Since our experiment's target is quite general, we decided to choose everyday products with neutral colors that could be used all genders. Therefore, a white water bottle, a black beanie and a grey backpack were chosen as products. Later, we created our survey on Qualtrics. The very first question on the survey was a default question which was shown to everyone regardless of their treatment/control group. We edited our treatment and control groups by adding a "sustainably sourced" green badge on one of the products in each product (water bottle, beanie and backpack) questions. The control group was shown the products with no badge, whereas the treatment group was shown products with the badge. Then, every participant was randomly assigned to either a treatment or control group by the randomizer feature on Qualtrics. They were asked to answer which product they would prefer for water bottle, beanie and backpack. Lastly, we performed t-test and regression analysis to see if the participants chose treatment product options significantly more than control in order to see if the green badge has an effect on conversion.

Participants

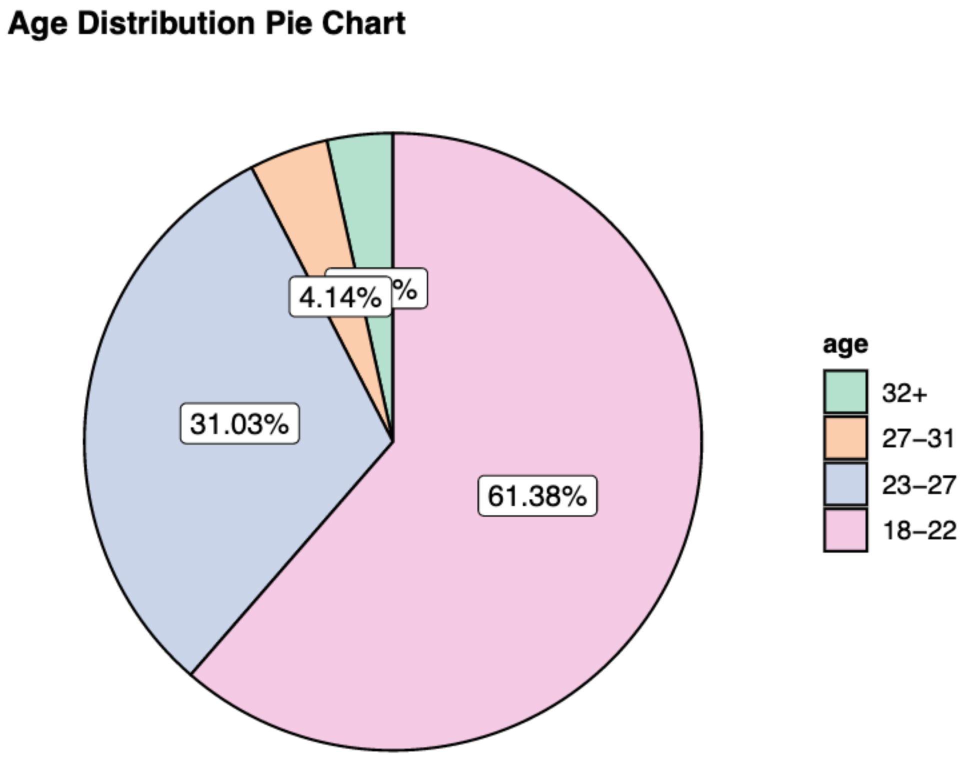
Since the goal of the experiment was to analyze whether companies that advertise their sustainability have a higher likelihood of leading to conversion of customers, we decided to target all demographics would participate for the experiment. Our initial method to recruit participants was by directly messaging our friends and families. We also posted our experiment on our Instagram stories in order to reach out to more people. People who were interested in participating were required to fill out a Qualtrics survey that started with the treatment questions and ended with the covariate questions asking them their age, gender and their undergraduate major. We closed the survey 4 days after distribution with a total of 145 responses. From the below charts, we can see the gender, age and undergraduate major distributions of the participants.

```r
pie <- ggpiestats(regression_table, 'gender',
           results.subtitle = F,
           slice.label = 'both', #show percentage together with counts
           perc.k = 2, #2 digits
           direction = 1,
           palette = 'Pastel2',
           title = 'Gender Distribution Pie Chart')
pie
```

Gender Distribution Pie Chart




```r
pie2 <- ggpiestats(regression_table, 'age',
           results.subtitle = F,
           slice.label = 'both', #show percentage together with counts
           perc.k = 2, #2 digits
           direction = 1,
           palette = 'Pastel2',
           title = 'Age Distribution Pie Chart')
pie2
```



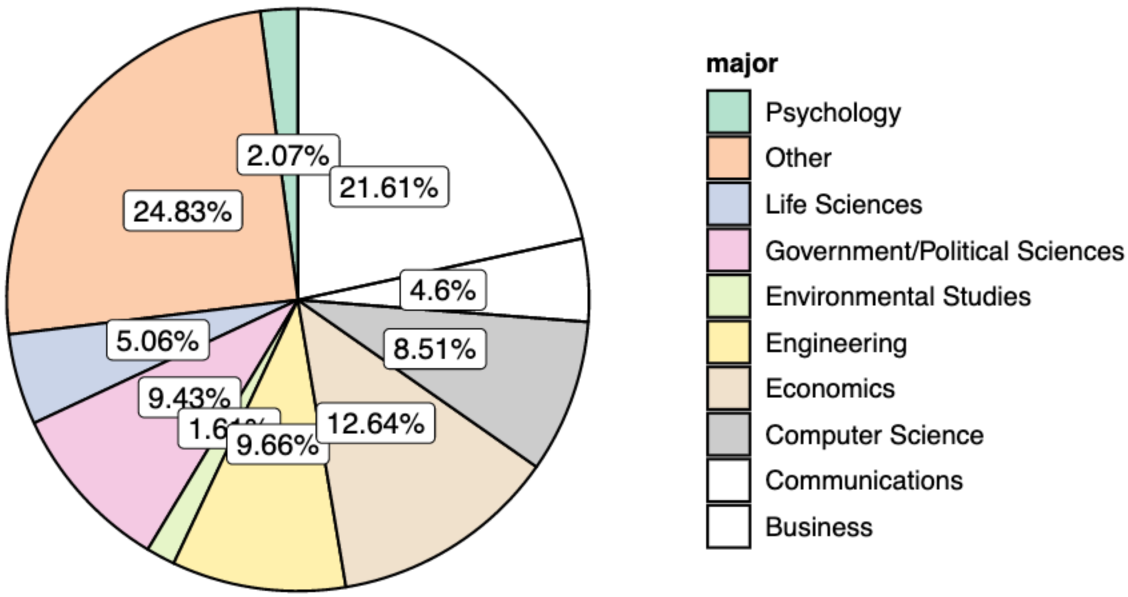

```
pie3 <- ggpiestats(regression_table, 'major',
          results.subtitle = F,
          slice.label = 'both', #show percentage together with counts
          perc.k = 2, #2 digits
          direction = 1,
          palette = 'Pastel2',
          title = 'Major Distribution Pie Chart')
```

```
## Warning: Number of labels is greater than default palette color count.
##  Try using another color `palette` (and/or `package`).

##
```

```
pie3
```

Major Distribution Pie Chart



Randomization

Since we did our experiment on Qualtrics, in order to decide which participants would be in the control group or treatment group, we used the Randomizer on Qualtrics. This feature on Qualtrics allowed us to randomly send participants to either control or treatment group. In our survey, for each product category, users were randomly placed into either the control or treatment group. For each category, the survey randomized between 1 control and 3 treatment questions, therefore, the ratio was 25% control and 75% treatment.

Pre-Experiment Randomization Check: Treatment

The prop test for treatment assignment indicates that we cannot reject the null that the randomization was properly done, as the p-value is greater than 0.05, and so we do not need to be concerned about the randomization process in this experiment.

```
prop.test(regression_table[treatment ==1, .N],435,0.75)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  regression_table[treatment == 1, .N] out of 435, null probability 0.75
## X-squared = 0.0375479, df = 1, p-value = 0.84635
## alternative hypothesis: true p is not equal to 0.75
## 95 percent confidence interval:
##  0.70064061 0.78460873
## sample estimates:
##          p
## 0.74482759
```

Pre-Experiment Randomization Check: Treatment Arm

The prop test for treatment arm assignment indicates that we cannot reject the null that the randomization was properly done, as the p-value is greater than 0.05, and so we do not need to be concerned about the randomization process in this experiment.

```r
prop.test(regression_table[arm ==0, .N],435,0.25)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  regression_table[arm == 0, .N] out of 435, null probability 0.25
## X-squared = 0.0375479, df = 1, p-value = 0.84635
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.21539127 0.29935939
## sample estimates:
##          p
## 0.25517241
```

```r
prop.test(regression_table[arm ==1, .N],435,0.25)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  regression_table[arm == 1, .N] out of 435, null probability 0.25
## X-squared = 0.0375479, df = 1, p-value = 0.84635
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.21539127 0.29935939
## sample estimates:
##          p
## 0.25517241
```

```r
prop.test(regression_table[arm ==2, .N],435,0.25)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  regression_table[arm == 2, .N] out of 435, null probability 0.25
## X-squared = 0.478927, df = 1, p-value = 0.48891
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.19602703 0.27772035
## sample estimates:
##          p
## 0.23448276
```

```r
prop.test(regression_table[arm ==3, .N],435,0.25)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  regression_table[arm == 3, .N] out of 435, null probability 0.25
## X-squared = 0.0375479, df = 1, p-value = 0.84635
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.21539127 0.29935939
## sample estimates:
##          p
## 0.25517241
```

Data Analysis

A. Analysis based on regression

With the data we have, we can run a number of regressions to analyze the behavior of our participants in varieties of ways. We begin with a simple regression analyzing the main effect we are researching – the effect for consumers' choice of being in the treatment group (seeing the product with green sustainable mark). Therefore, we ran the following regression with corresponding results.

A. i) Regression of main effect on outcomes

```
reg1 <- feols(outcome ~ treatment, data = regression_table, se = 'hetero')
reg1 %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat","P-Value"), digits =
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.000	0.000	-8.446	0
treatment	0.481	0.028	17.305	0

Based on the results of regression one, it is observed that being treated, i.e. seeing a sustainable mark on a product, increased consumer preference for that product. The coefficient is 0.481 and is found to be statistically significant. The std error is 0.028.

We also wondered what would happen if we add our covariates into our regression, will the result be quite different?

A. ii) Regression of main effect plus covariates on outcome

```
# add covariate - gender
reg2 <- feols(outcome ~ treatment + gender, data = regression_table, se = 'hetero')
reg2 %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat","P-Value"), digits =
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	-0.007	0.012	-0.546	0.586
treatment	0.496	0.028	17.421	0.000
genderMale	0.010	0.049	0.205	0.837
genderNon-binary/non-conforming	-0.489	0.032	-15.473	0.000
genderOther	-0.489	0.032	-15.473	0.000
genderPrefer not to say	-0.139	0.146	-0.954	0.340
genderTransgender	0.309	0.113	2.730	0.007

```
# add covariate - age
reg3 <- feols(outcome ~ treatment + age, data = regression_table, se = 'hetero')
reg3 %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat","P-Value"), digits =
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	-0.001	0.015	-0.041	0.967
treatment	0.478	0.028	16.912	0.000
age23-27	-0.027	0.046	-0.591	0.555
age27-31	0.102	0.110	0.932	0.352
age32+	0.218	0.108	2.023	0.044

```
# add covariate - major
reg4 <- feols(outcome ~ treatment + major, data = regression_table, se = 'hetero')
reg4 %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat","P-Value"), digits =
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.109	0.044	2.480	0.014
treatment	0.464	0.029	15.734	0.000
majorCommunications	0.213	0.074	2.887	0.004
majorComputer Science	-0.123	0.085	-1.453	0.147
majorEconomics	-0.194	0.070	-2.753	0.006
majorEngineering	-0.116	0.083	-1.397	0.163
majorEnvironmental Studies	0.351	0.086	4.084	0.000
majorGovernment/Political Sciences	-0.043	0.085	-0.508	0.612
majorLife Sciences	-0.289	0.089	-3.246	0.001
majorOther	-0.194	0.061	-3.175	0.002
majorPsychology	0.094	0.166	0.566	0.571

In order to try and reduce standard error, Regressions 2,3,and 4 attempted different combinations of covariate variables. We regressed for main outcome adding covariate of 'gender', 'age', and 'major.

After individually testing the impact of adding 'gender', 'age', and 'major' as separate covariates in regressions 2,3, and 4 respectively, it was observed that neither the coefficient, nor the standard error changed substantially. What this might indicate is that these variables did not account for a large amount of variance in this study and so did not improve precision of the initial results.

B. Analysis based on t-test

B. i) Outcome and Treatment t-test

The below t-test compares the outcome based on whether there was a sustainable badge or not. The results corroborate Regression 1, and indicate that the true difference in means is not equal to 0, rejecting the null hypothesis.

```r
# t test outcomes by treatment
t.test(regression_table[treatment == 0, outcome],
       regression_table[treatment == 1, outcome])
```

```
##
##  Welch Two Sample t-test
##
## data:  regression_table[treatment == 0, outcome] and regression_table[treatment == 1, outcome]
## t = -17.3184, df = 323, p-value < 2.22e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53617667 -0.42678629
## sample estimates:
##  mean of x  mean of y
## 0.00000000 0.48148148
```

B. ii) Comparing Treatment Arms 1 and 2

The t-test results comparing treatment arms 1 and 2 indicates that true difference in means is not equal to 0, and so the null hypothesis is rejected. The mean of treatment arm 1 is also observed to be higher than treatment arm 2.

```r
# t test outcomes by treatment
t.test(regression_table[arm == 1, outcome],
       regression_table[arm == 2, outcome])
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  regression_table[arm == 1, outcome] and regression_table[arm == 2, outcome]
## t = 3.5129, df = 204.883, p-value = 0.00054528
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.10184072 0.36238822
## sample estimates:
##  mean of x  mean of y
## 0.70270270 0.47058824
```

B. iii) Comparing Treatment Arms 2 and 3

The t-test results comparing treatment arms 2 and 3 indicates that true difference in means is not equal to 0, and so the null hypothesis is rejected. The mean of treatment arm 2 is also observed to be higher than treatment arm 3.

```r
# t test outcomes by treatment
t.test(regression_table[arm == 2, outcome],
       regression_table[arm == 3, outcome])
```

```
##
##  Welch Two Sample t-test
##
## data:  regression_table[arm == 2, outcome] and regression_table[arm == 3, outcome]
## t = 3.06926, df = 202.806, p-value = 0.0024393
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.071631332 0.329004598
## sample estimates:
##  mean of x  mean of y
## 0.47058824 0.27027027
```

B. iv) Comparing Treatment Arms 1 and 3

The t-test results comparing treatment arms 1 and 3 indicates that true difference in means is not equal to 0, and so the null hypothesis is rejected. The mean of treatment arm 1 is also observed to be higher than treatment arm 3.

```r
# t test outcomes by treatment
t.test(regression_table[arm == 1, outcome],
       regression_table[arm == 3, outcome])
```

```
##
##  Welch Two Sample t-test
##
## data:  regression_table[arm == 1, outcome] and regression_table[arm == 3, outcome]
## t = 7.1167, df = 219.818, p-value = 1.545e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.31267979 0.55218507
## sample estimates:
##  mean of x  mean of y
## 0.70270270 0.27027027
```

Limitations

An ideal experiment for our project would be to have a more diverse sample of people taking the survey. When sending out the survey, we mainly sent it out to the people that are around us, who happen to be around the same age range and similar educational backgrounds. Since we did not do blocking on gender and age range, we found that we have an disproportionate amount of females to males and people aged 18-22 over all other ages. The percentage of females who took the survey made up 66% of the sample population and those aged 18-22 made up 62% of the sample population. In an ideal situation, we would have a better representation of the different genders and ages in our experiment. This is a threat to external validity because it does not accurately represent the true population, since some groups are more represented than others in our sample data set.

Additionally, the scope of the project is limited because with this survey we are not measuring purchase behavior but consumer preferences. In our survey we presented our surveyees with different products and randomized the different products in each category with a "sustainably sourced" tag and they had to choose which they would prefer over the other presented products. This doesn't equate to whether consumers will buy sustainably sourced products, it just indicates consumer preferences towards sustainability. It would be interesting to see how this can be transformed to actual purchases to better understand if people tend to buy sustainably sourced products more than products that aren't sustainably sourced.

Another limitation that we have in our project is that in the three categories of products we included, we included two wearable items (beanie and backpack) and one non-wearable item (water bottle). This could have had an impact on what people prefer to be sustainably sourced or not, as the treatment arm with the water bottle had the greatest coefficient. Other factors such as style could also dictate whether someone chooses one product over another regardless of whether it's sustainable, especially for wearable items over non-wearable items.

Finally, another possible source of bias is self-selection. Participants of the experiment were asked to complete this survey and this could have created some extra noise in the experiment.

Conclusion

In this experiment, we wanted to test whether consumers significantly prefer sustainable products over non-sustainable products. The variable that we wanted to measure was whether the people who took the survey chose a sustainably sourced product over the other options presented. In designing the experiment we created a survey consisting of three categories of products with three options each and the surveyee would have to choose one option from the three. Each participant was randomly assigned to the different treatment arms for each of the three categories and it was equally split as to how many were in the control group and how many were in the treatment groups (each arm had the same number of participants).

We did a pre-experiment randomization check on both treatment vs. control group as well as on each of the treatment arms. From the pre-experiment randomization check we found out that we randomized well. After that, we ran multiple regression analyses and found that the treatment was statistically significant. We also found out that the treatment group for the water bottle category (arm 1) was more statistically significant than the other two categories, which were a hat and a backpack. We can conclude that people might be more likely to prefer sustainably sourced for non-wearable items or that wearable items are more complicated to determine since people might prefer style over sustainability; however, more experiments should be run to determine this.

Despite the statistically significant positive effect of the treatment, we cannot imply causality due to non-interference and excludability. Because the survey was distributed between our colleagues they may have interacted while completing the survey which violates the non-interference assumption. The excludability assumption may have been violated because of the nature of the products, they were not all exactly the same; therefore, this may have affected consumer preference.

Going forward, we can improve our experiment by sampling a larger population size and perform blocking on certain categories so that we have a better representation of the true population. Also, we can expand on this experiment so that we also test purchase behavior in addition to consumer preferences.

Bibiliography

Rume, T., & Islam, S. M. D.-U. (2020, September 17). Environmental effects of COVID-19 pandemic and potential strategies of Sustainability. National Center for Biotechnology Information. Retrieved December 7, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7498239/.

Whelan, T., & Kronthal-Sacco, R. (2019, June 19). Research: Actually, consumers do buy sustainable products. Harvard Business Review. Retrieved December 7, 2021, from https://hbr.org/2019/06/research-actually-consumers-do-buy-sustainable-products.

Appendix 1 – Table of Survey (except for their choices towards products)

Boston University

Default Question Block

Brief Overview

Thank you for taking the time to participate in this survey!

We are a group of students looking to analyze consumer preference within certain categories of products. Throughout the survey you will be presented with various products and you will be prompted to select one. Please select your preference.

This survey should take no longer than 5 minutes.

Covariate Questions

What gender do you identify with?

- Female
- Male
- Transgender
- Non-binary/non-conforming
- Other
- Prefer not to say

How old are you?

- 18-22
- 23-27
- 27-31
- 32+

What was your undergraduate concentration?

- Computer Science
- Communications
- Government/Political Sciences
- Business
- Economics
- Psychology
- Engineering
- Life Sciences
- Environmental Studies
- Other

Appendix 2 – Pictures of Products (Those Green Badge shows up randomly)

Product 1 – Water Bottles

Product 2 – Beanies

Product 3 – Backpacks