

BA 830 Project

Team 10

4/26/2021

Team members: Paula Demacker, Yuchen Feng, Leah Fowlkes, Ryan Veytsman, Xinping Yu, Yuzhe Zheng

Introduction

Our goal is to understand if the color of the plate that food is served on affects whether people want to eat the food. Then we can give some recommendation to some restaurant to help them to make a better advertisement to attract more people.

We created two surveys and used simply randomization to randomize by person level to put each person into treatment group or control group with different surveys. For each survey, we have basic demographic information problems such as gender, country and age group to see if any these information may also affect on the plate that they chose. We also have 10 questions in each survey relates to our experiment. In these questions, We asked people to select based on the photos that which food looks tastier. We have ten different foods in total and for each question we display the same food but different food pictures into different colors of plate. We randomized for each question that the red plate will on the right or left and set the treatment survey reverse the plate color order for each question. And we will test people will more likely to choose the food based on the plate color, food itself or the position of the answer.

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.5      v dplyr   1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##   expand, pack, unpack

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
```

```

##      hour, isoweek, mday, minute, month, quarter, second, wday, week,
##      yday, year

## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union

##
## Please cite as:
##
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

#import the file
food <- fread('BA830-Project.csv')
food = subset(food, select = -c(outcome_question)) #to get rid of extra column
food$outcome_plate.dummy <- ifelse(food$outcome_plate==0,1,0) #to recode into new column properly
food$outcome_plate <- food$outcome_plate.dummy #to properly assign outcome_plate
food = subset(food, select = -c(outcome_plate.dummy)) #to rid of new extra column

head(food)

##      Person_id  Age Currently living Gender outcome_plate outcome_food
## 1:           1 46-55           Europe  Woman             0             1
## 2:           1 46-55           Europe  Woman             0             1
## 3:           1 46-55           Europe  Woman             0             0
## 4:           1 46-55           Europe  Woman             1             0
## 5:           1 46-55           Europe  Woman             0             1
## 6:           1 46-55           Europe  Woman             1             1
##      food_name Treatment
## 1:      Burger          0
## 2:       Taco          0
## 3: Fried Chicken          0
## 4:      Salmon          0
## 5:       Pasta          0
## 6:       Steak          0

#data cleaning
#gender
food$gender_group[food$Gender == 'Man'] <- 1
food$gender_group[food$Gender == 'Woman'] <- 0
#age
food$age_group[food$Age == '16-25'] <- 1
food$age_group[food$Age == '26-35'] <- 2
food$age_group[food$Age == '36-45'] <- 3
food$age_group[food$Age == '46-55'] <- 4
food$age_group[food$Age == '56-65'] <- 5
#country
food$country_group[food$'Currently living' == 'Europe'] <- 1
food$country_group[food$'Currently living' == 'North America/Central America'] <- 2
food$country_group[food$'Currently living' == 'Asia/Middle east'] <- 3
food$country_group[food$'Currently living' == 'Australia/New Zealand'] <- 4

#proportion test
prop.test(nrow(food[Treatment == 1])/10, nrow(food)/10, 0.5)

##

```

```
## 1-sample proportions test with continuity correction
##
## data: nrow(food[Treatment == 1])/10 out of nrow(food)/10, null probability 0.5
## X-squared = 1, df = 1, p-value = 0.3
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.383 0.534
## sample estimates:
##      p
## 0.458
```

Our group are using simple randomization to select our treatment group and control group. By using the proportion test we can see the p value here is 0.3 which is larger than 0.05, so we fail to reject the null hypothesis and prove that our experiment of treatment and control group was well randomized.

Randomization / balance check for pre-experiment variables between the treatment and control group.

```
reg.age <- lm(age_group ~ Treatment, data = food)
summary(reg.age)
```

```
##
## Call:
## lm(formula = age_group ~ Treatment, data = food)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.563 -0.563 -0.543  0.437  3.457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5625     0.0316   49.41  <2e-16 ***
## Treatment    -0.0193     0.0467   -0.41    0.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.98 on 1768 degrees of freedom
## Multiple R-squared:  9.63e-05, Adjusted R-squared:  -0.000469
## F-statistic: 0.17 on 1 and 1768 DF, p-value: 0.68
reg.gender <- lm(gender_group ~ Treatment, data = food)
summary(reg.gender)
```

```
##
## Call:
## lm(formula = gender_group ~ Treatment, data = food)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.532 -0.532  0.468  0.468  0.506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5319     0.0163   32.63  <2e-16 ***
```

```
## Treatment      -0.0381      0.0240     -1.59      0.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5 on 1748 degrees of freedom
## (20 observations deleted due to missingness)
## Multiple R-squared:  0.00144,    Adjusted R-squared:  0.000873
## F-statistic: 2.53 on 1 and 1748 DF,  p-value: 0.112

reg.country <- lm(country_group ~ Treatment, data = food)
summary(reg.country)
```

```
##
## Call:
## lm(formula = country_group ~ Treatment, data = food)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.427 -0.427  0.573  0.573  1.667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.4271     0.0245   99.25  <2e-16 ***
## Treatment    -0.0938     0.0361   -2.59   0.0096 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.758 on 1768 degrees of freedom
## Multiple R-squared:  0.00379,    Adjusted R-squared:  0.00323
## F-statistic: 6.73 on 1 and 1768 DF,  p-value: 0.00958

reg.plate <- lm(outcome_plate ~ Treatment, data = food)
summary(reg.plate)
```

```
##
## Call:
## lm(formula = outcome_plate ~ Treatment, data = food)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.469 -0.469 -0.295  0.531  0.705
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4688     0.0155   30.25  < 2e-16 ***
## Treatment    -0.1737     0.0229   -7.58  5.4e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.48 on 1768 degrees of freedom
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0309
## F-statistic: 57.5 on 1 and 1768 DF,  p-value: 5.44e-14

reg.question <- lm(outcome_food ~ Treatment, data = food)
summary(reg.question)
```

```
##
## Call:
## lm(formula = outcome_food ~ Treatment, data = food)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.486 -0.486 -0.404  0.513  0.596
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4865     0.0160   30.4 < 2e-16 ***
## Treatment    -0.0828     0.0237   -3.5  0.00048 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.496 on 1768 degrees of freedom
## Multiple R-squared:  0.00687,    Adjusted R-squared:  0.00631
## F-statistic: 12.2 on 1 and 1768 DF,  p-value: 0.000481
```

By running regression for all the pre-experiment variables of regression that we select based on our survey, we find out that gender_group and age_group gave us a big p value bigger than 0.05 which is equal to 0.11 and 0.68. In this case, we are not able to reject the null hypothesis. It means that gender_group and age_group are well randomized between treatment group and control group. However, the p value for country_group is smaller than 0.05 which is 0.0096. In this case, we are able to reject the null hypothesis. It means that country_group is not well randomized between treatment group and control group. The reason for that is because we are using simple randomization, and the sample size for some country is not big enough. For example, we might only have one observation from Country A, and based on our simple randomization, it will make our data be bias.

Regression

```
together <- feols(c(outcome_plate, outcome_food) ~ Treatment,
                  data = food, se = 'white')
etable(together)
```

```
##              model 1              model 2
## Dependent Var.: outcome_plate outcome_food
##
## (Intercept)    0.4687*** (0.0161)  0.4865*** (0.0161)
## Treatment      -0.1737*** (0.0227) -0.0828*** (0.0236)
## -----
## S.E. type      Heteroskedast.-rob. Heteroskedast.-rob.
## Observations              1,770              1,770
## R2              0.03150              0.00687
## Adj. R2         0.03095              0.00631
```

In the regression of model 1 we looked at the outcome_plate as the outcome variable first. We get an intercept of 0.4687, which represents the mean of the outcome of our control group. The estimate for the treatment effect is -0.1737 and has a standard error of 0.0227. Therefore, we have a negative treatment effect which indicates a negative relationship between Treatment and outcome_plate.

We know from the stars that this result is statistically significant. This means that people in the treatment group are more likely to choose the white plate.

For model 2 we used outcome_food as our outcome variable. The intercept is 0.4865. Our estimated treatment effect is -0.0828 and has a standard error of 0.0236. This indicates that we have a negative relationship

between Treatment and outcome_food. These results are also statistically significant. It still indicates that people in the treatment group are more likely to choose the food on the red plate. The negative treatment effect shows us that people in the treatment group are more likely to choose the plate that is positioned on the right.

```
fixed_effects <- feols(outcome_plate ~ Treatment
  | outcome_food + food_name, data = food,
  se = 'white')
etable(fixed_effects)
```

```
##                               fixed_effects
## Dependent Var.:               outcome_plate
##
## Treatment                    -0.1788*** (0.0230)
## Fixed-Effects:  -----
## outcome_food                  Yes
## food_name                     Yes
## -----
## S.E. type                     Heteroskedast.-rob.
## Observations                  1,770
## R2                            0.04394
## Within R2                     0.03349
```

For the regression above we added two fixed effects: outcome_food and food_name to see if the food has an impact on the outcome. The outcome variable is outcome_plate and has an estimated treatment effect of -0.1788 with an standard error of 0.0230, which is really similar to the regression above without the fixed effects. The results of this regression are also statistically significant.

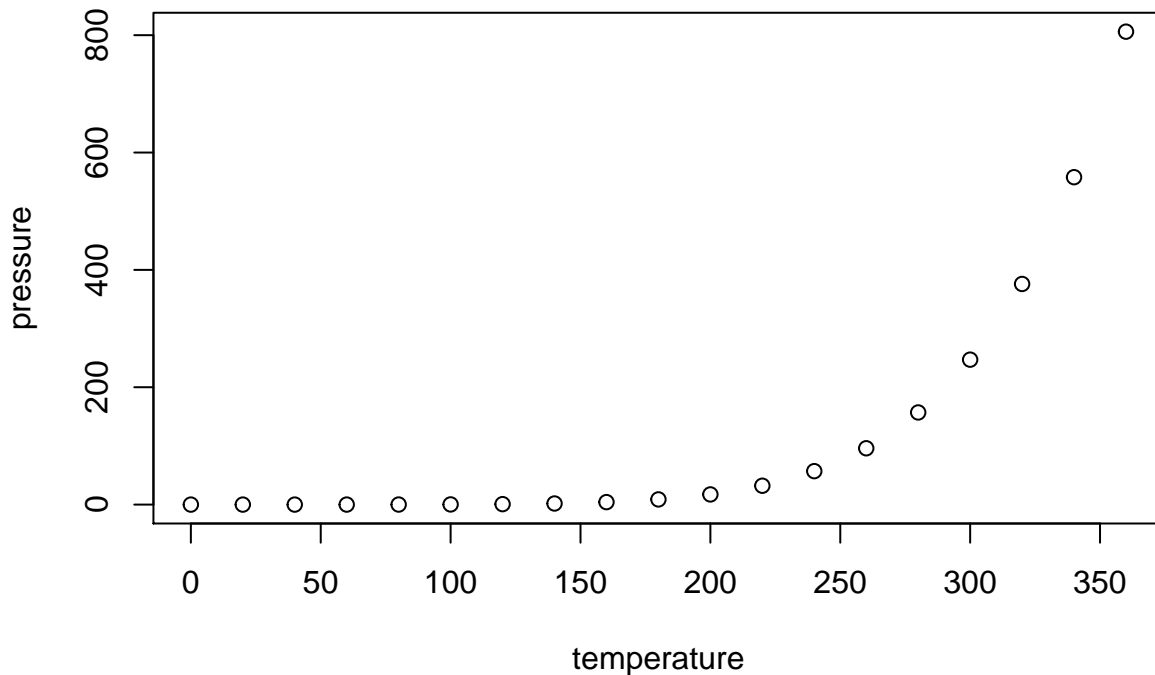
```
reg.burger <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Burger'])
reg.taco <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Taco'])
reg.friedchicken <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Fried Chicken'])
reg.salmon <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Salmon'])
reg.pasta <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Pasta'])
reg.steak <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Steak'])
reg.dumpling <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Dumpling'])
reg.sushi <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Sushi'])
reg.friedrice <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Fried Rice'])
reg.salad <- feols(outcome_plate ~ Treatment, data = food[food_name == 'Salad'])
etable(reg.burger, reg.taco, reg.friedchicken, reg.salmon, reg.pasta, reg.steak, reg.dumpling, reg.sushi)
```

```
##                               reg.burger    reg.taco    reg.friedchicken
## Dependent Var.:               outcome_plate    outcome_plate    outcome_plate
##
## (Intercept)                   0.6146*** (0.0474) 0.4063*** (0.0463) 0.4063*** (0.0503)
## Treatment                     -0.3677*** (0.0701) -0.2087** (0.0684)  -0.0112 (0.0744)
## -----
## S.E. type                     Standard          Standard          Standard
## Observations                  177                177                177
## R2                            0.13577            0.05049            0.00013
## Adj. R2                       0.13084            0.04506            -0.00558
##                               reg.salmon    reg.pasta    reg.steak
## Dependent Var.:               outcome_plate    outcome_plate    outcome_plate
##
## (Intercept)                   0.5938*** (0.0413) 0.3438*** (0.0493) 0.3438*** (0.0486)
## Treatment                     -0.5197*** (0.0611) 0.0390 (0.0728) -0.0104 (0.0718)
## -----
```

```
## S.E. type          Standard          Standard          Standard
## Observations      177              177              177
## R2                0.29240          0.00163          0.00012
## Adj. R2           0.28836          -0.00407         -0.00559
##                   reg.dumpling      reg.sushi       reg.friedrice
## Dependent Var.:   outcome_plate    outcome_plate    outcome_plate
##
## (Intercept)       0.6563*** (0.0420) 0.3021*** (0.0491) 0.6771*** (0.0410)
## Treatment         -0.5451*** (0.0621) 0.1917** (0.0726) -0.5783*** (0.0606)
## -----
## S.E. type          Standard          Standard          Standard
## Observations      177              177              177
## R2                0.30567          0.03836          0.34256
## Adj. R2           0.30170          0.03287          0.33880
##                   reg.salad
## Dependent Var.:   outcome_plate
##
## (Intercept)       0.3438*** (0.0493)
## Treatment         0.2735*** (0.0728)
## -----
## S.E. type          Standard
## Observations      177
## R2                0.07457
## Adj. R2           0.06928
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Heterogeneous Treatment Effects

We will use the `feols` function to compute the heterogeneous treatment effect of being shown the red plate by whether the person saw the food with the red plate on the left (first option) or right (second option) of the screen.

```
reg_het1 <- feols(outcome_plate ~ Treatment*outcome_food, data = food, se = 'white')
etable(reg_het1)
```

```
##                                reg_het1
## Dependent Var.:                outcome_plate
##
## (Intercept)                    0.5700*** (0.0223)
## Treatment                      -0.3257*** (0.0297)
## outcome_food                   -0.2081*** (0.0315)
## Treatment x outcome_food       0.3338*** (0.0457)
## -----
## S.E. type                      Heteroskedast.-rob.
## Observations                    1,770
## R2                             0.06350
## Adj. R2                        0.06191
```

Explanation of the Results

If a person only takes the treatment survey, they were less likely to pick the red plate 32.6% of the time than if they took the control survey. We also see if the food was on the left or listed as the first option, they were 20.8% less likely to pick the first option than the second. For the heterogeneous treatment effect, if the person took the treatment survey and chose the first option of the survey, they were 33.4% more likely to pick the red plate. All of these results are at a high significance level so we can believe these trends stand for 99% of the time. The combined treatment effect led more people to choose the red plate than the white plate. We also see the treatment and the positioning of the red plate on their own led to more people choosing the white plate over the red.

We will use the `feols` function again to compute the heterogeneous treatment effect of being shown the red plate by whether the person saw a certain type or category of food.

```
reg_het2 <- feols(outcome_plate ~ Treatment*food_name, data = food, se = 'white')
etable(reg_het2)
```

```
##                                reg_het2
## Dependent Var.:                outcome_plate
##
## (Intercept)                    0.6146*** (0.0500)
## Treatment                      -0.3677*** (0.0694)
## food_nameDumpling              0.0417 (0.0698)
## food_nameFriedChicken         -0.2083** (0.0710)
## food_nameFriedRice            0.0625 (0.0693)
## food_namePasta                -0.2708*** (0.0698)
## food_nameSalad                -0.2708*** (0.0698)
## food_nameSalmon               -0.0208 (0.0710)
## food_nameSteak                -0.2708*** (0.0698)
## food_nameSushi                -0.3125*** (0.0687)
## food_nameTaco                 -0.2083** (0.0710)
## Treatment x food_nameDumpling -0.1775. (0.0918)
```



```

## Treatment x food_nameFriedChicken  0.3565*** (0.1017)
## Treatment x food_nameFriedRice      -0.2106* (0.0907)
## Treatment x food_namePasta          0.4066*** (0.1007)
## Treatment x food_nameSalad          0.6412*** (0.1007)
## Treatment x food_nameSalmon         -0.1520. (0.0906)
## Treatment x food_nameSteak          0.3573*** (0.0998)
## Treatment x food_nameSushi          0.5594*** (0.1008)
## Treatment x food_nameTaco           0.1590 (0.0966)
## -----
## S.E. type                          Heteroskedast.-rob.
## Observations                        1,770
## R2                                  0.13353
## Adj. R2                             0.12413

```

Explanation of the Results

For this regression, taking the treatment survey led people to choose the red plate 36.8% of the time less than if they took the control survey. There were 10 food categories: Burger, Dumpling, Fried Chicken, Fried Rice, Pasta, Salad, Salmon, Steak, Sushi, and Taco. Their respective treatment effects, with Burger standing as the 0 or the variable left out of the regression, were (0.0417, -0.2083, **0.0625**, **-0.2708**, *-0.2708*, **-0.0208**, *-0.3125*, **-0.2083**). Some of the food name effects were not at a high significance level so we will not report those. We see if Fried chicken or Taco was displayed, they were 21% less likely to choose the red plate option. For Pasta, Salad, and Steak, people tended to pick the red plate option for these meals 27% of the time less than the other option. People were 31% less likely to choose the red plate option if the food category was sushi. These results are in comparison to the burger on a red plate. In general, the burger, dumpling and Salmon tended to be preferred on the red plate than the other dishes. For the heterogeneous effect, if people took the treatment survey and were shown a specific food category, the treatment effects are as follows: (-0.1775, 0.3565, **-0.2106**, **0.4066**, *0.6412*, **-0.1520**, **0.3573**, *0.5594***, 0.1590). The fried chicken images led people to choose the red plate less often by 36%. Fried rice caused a 21% lower chance people would pick the red plate option. The pasta category created a 21% lower chance the red plate option would be picked, while salad, steak, and sushi led to their red plate option being more likely chosen 64%, 36%, and 56% of the time.

Limitations

Sample Size

One limitation to our regressions would be our sample size. Our sample only consisted of 186 survey results. This is not nearly enough observations to determine a trend on a larger scale so our statistical power is very low. Our probability of rejecting the null when there is a true treatment effect of some size is substantially low.