

W241 Final Project: Subliminal Investigation

Amit Karandikar

12/10/2020

```
d <- fread('./data/survey-responses.csv', header = TRUE)

# Questions 8 - 17 are the quiz
setnames(
  x = d,
  old = c("SC0", "Q_RecaptchaScore", "Q2", "Q3", "Q4", "Q6",
          "Q18", "Q19", "Q20", "Q21", "Q26"),
  new = c("score", "bot_response", "interest", "experience",
          "sub_treatment", "sub_control", "used_ext",
          "gender", "age", "education",
          "country")
)
# make all columns lowercase
setnames(d, tolower(names(d[1:ncol(d)])))

# Remove invalid data if experience has NA values - no randomization occurred
d <- d[!is.na(experience)]

# count the no of questions answered
d[, questions_attempted := (10 - (is.na(q8) + is.na(q9) + + is.na(q10) +
                                is.na(q11) + is.na(q12) + is.na(q13) +
                                is.na(q14) + is.na(q15) + is.na(q16) + is.na(q17)))]

# convert posix timestamps into dates
d[, `:=` (startdate = as.POSIXct(startdate), enddate = as.POSIXct(enddate))]

# Remove pilot group as well as those uninterested in Data Science (Question 2)
d <- d[startdate > '2021-11-12 19:00' & interest == 1 ]

# Remove Bot responses (scores will be less than 0.5)
# d <- d[bot_response >= 0.5]
paste('Total valid responses: ', nrow(d))

## [1] "Total valid responses: 518"

# mark treatment and control groups
d[, subliminal := ifelse(test = d[, is.na(sub_treatment)], 0, 1)]
d[, experience_ind := ifelse(test = d[, experience > 2], 1, 0)]
paste('No of people in treatment', nrow(d[subliminal == 1]))

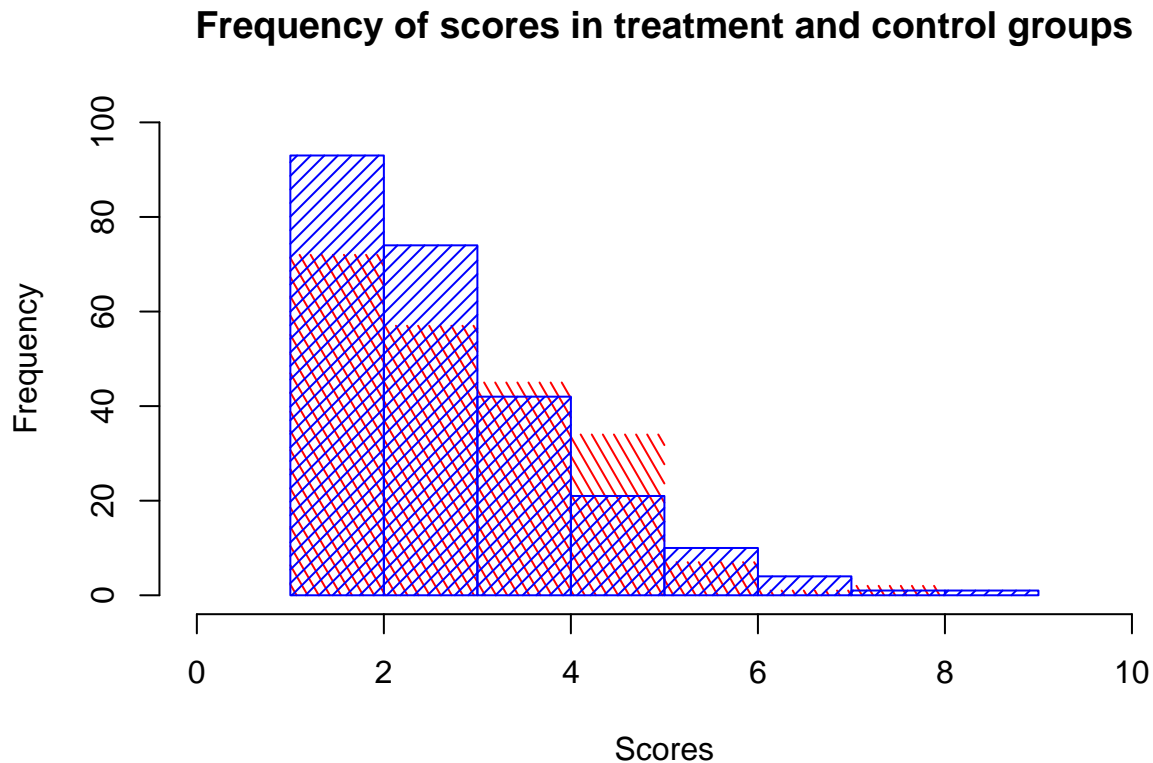
## [1] "No of people in treatment 273"
```

```
paste('No of people in control', nrow(d[subliminal==0]))
```

```
## [1] "No of people in control 245"
```

```
hist(d[which(d$subliminal == 0) ,]$score,
     xlim = c(0,10),
     ylim = c(0, 100), col = "red",
     border = F,
     main = "Frequency of scores in treatment and control groups",
     xlab = "Scores", density=20, angle = 120)
```

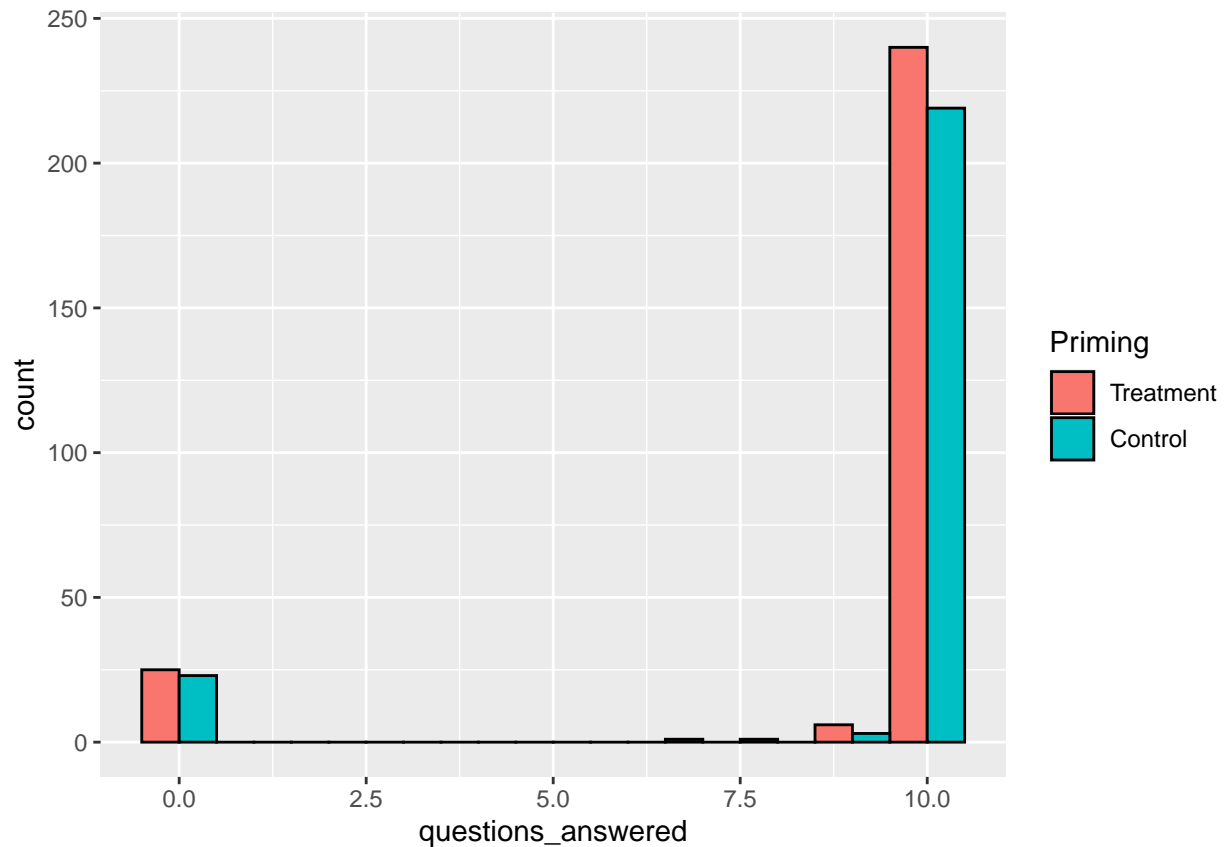
```
hist(d[which(d$subliminal == 1),]$score,
     xlim = c(0,10),
     ylim = c(0, 100), col= "blue",
     add=T, density = 20)
```



```
require(ggplot2)
```

```
dplot <- rbind(data.frame(ev=rgb(1,0,0), questions_answered = d[which(d$subliminal == 0) ,]$questions_a
                  data.frame(ev=rgb(0,0,1), questions_answered = d[which(d$subliminal == 1) ,]$questions_a
```

```
ggplot(dplot, aes(x=questions_answered, fill=ev, )) +
  geom_histogram(binwidth=1, colour="black", position="dodge") +
  theme(legend.position = "right") + scale_fill_discrete(name="Priming", labels=c('Treatment', 'Control'))
```



Calculate the average treatment effect

```
ate_model <- d[,lm(score ~ subliminal)]
summary(ate_model, vcov=vcovHC)
```

```
##
## Call:
## lm(formula = score ~ subliminal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2523 -1.0732 -0.0732  0.9268  5.9268
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.25229    0.09732   33.42  <2e-16 ***
## subliminal  -0.17912    0.13365   -1.34    0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.437 on 462 degrees of freedom
## (54 observations deleted due to missingness)
## Multiple R-squared:  0.003873,    Adjusted R-squared:  0.001717
```

```
## F-statistic: 1.796 on 1 and 462 DF, p-value: 0.1808
```

Effect of blocking

```
# indicator variable instead of factor , novice 0, expert 1
ate_blocking <- d[,lm(score ~ subliminal + experience_ind)]
summary(ate_blocking)

##
## Call:
## lm(formula = score ~ subliminal + experience_ind)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2892 -1.1084 -0.1084  0.8916  5.8916
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.2892     0.1042  31.574  <2e-16 ***
## subliminal     -0.1808     0.1337  -1.352    0.177
## experience_ind -0.1576     0.1589  -0.992    0.322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.437 on 461 degrees of freedom
## (54 observations deleted due to missingness)
## Multiple R-squared:  0.005995, Adjusted R-squared:  0.001683
## F-statistic:  1.39 on 2 and 461 DF, p-value: 0.2501
```

For this model, the estimated difference in treatment means is the average of the within-block differences. Individuals within each block are intended to be more similar based on their experience in working in the field of Data Science, and as a result, the variability of the within-block differences is expected to be smaller than the variability between two units randomly chosen from all units in the population.

All Covariates Model

```
ate_all_cov_model <- d[,lm(score ~ subliminal + gender
+ age + education + country
+ experience_ind
+ subliminal*education + subliminal*gender)]
summary(ate_all_cov_model)

##
## Call:
## lm(formula = score ~ subliminal + gender + age + education +
##      country + experience_ind + subliminal * education + subliminal *
##      gender)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6547 -1.0575 -0.0715  0.7069  5.6476
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.926338   0.534531   3.604 0.000349 ***
## subliminal       1.048608   0.633002   1.657 0.098311 .
## gender           0.057863   0.216119   0.268 0.789024
## age              0.029668   0.068372   0.434 0.664555
## education        0.262645   0.108257   2.426 0.015657 *
## country          0.002530   0.001013   2.497 0.012882 *
## experience_ind   -0.342510   0.174938  -1.958 0.050866 .
## subliminal:education -0.266008   0.140882  -1.888 0.059654 .
## subliminal:gender  -0.291712   0.278227  -1.048 0.294991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.394 on 446 degrees of freedom
## (63 observations deleted due to missingness)
## Multiple R-squared:  0.04994,    Adjusted R-squared:  0.03289
## F-statistic:  2.93 on 8 and 446 DF,  p-value: 0.003365
```

```
stargazer::stargazer(ate_model, ate_blocking, type = 'text',
                      column.labels = c('Model 1', 'Model 2'),
                      dep.var.labels = c('Quiz Score'),
                      covariate.labels = c('subliminal priming', 'experience'))
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               Quiz Score
##                               Model 1      Model 2
##                               (1)         (2)
## -----
## subliminal priming          -0.179        -0.181
##                               (0.134)      (0.134)
##
## experience                   -0.158
##                               (0.159)
##
## Constant                    3.252***      3.289***
##                               (0.097)      (0.104)
## -----
## Observations                464           464
## R2                          0.004         0.006
## Adjusted R2                 0.002         0.002
## Residual Std. Error  1.437 (df = 462)    1.437 (df = 461)
## F Statistic          1.796 (df = 1; 462)  1.390 (df = 2; 461)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
stargazer::stargazer(ate_model, ate_blocking, ate_all_cov_model, type = 'text')
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               score
##                               (1)      (2)      (3)
## -----
## subliminal      -0.179      -0.181      1.049*
##                  (0.134)      (0.134)      (0.633)
##
## gender                      0.058
##                             (0.216)
##
## age                      0.030
##                             (0.068)
##
## education            0.263**
##                      (0.108)
##
## country              0.003**
##                      (0.001)
##
## experience_ind      -0.158      -0.343*
##                      (0.159)      (0.175)
##
## subliminal:education -0.266*
##                      (0.141)
##
## subliminal:gender   -0.292
##                      (0.278)
##
## Constant           3.252***      3.289***      1.926***
##                     (0.097)      (0.104)      (0.535)
## -----
## Observations           464           464           455
## R2                     0.004           0.006           0.050
## Adjusted R2            0.002           0.002           0.033
## Residual Std. Error   1.437 (df = 462)   1.437 (df = 461)   1.394 (df = 446)
## F Statistic           1.796 (df = 1; 462) 1.390 (df = 2; 461) 2.930*** (df = 8; 446)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```