Computer Assignment 2b Costello

Rebecca Costello

9/9/2019

Rebecca Costello - Ties van der Veen - Niels van Opstal

Task: Estimate and interpret the effect of a treatment in another randomized trial.

Research Question: Are applications from guests with distinctively African American names less likely to be accepted relative to identical guests with distinctively white names.

```
knit_print.data.frame <- lemon_print
```

```
st_options(plain.ascii = FALSE, style = "rmarkdown")
st_css()
```

```
## <style type="text/css">
##  img {   background-color: transparent;   border: 0; }  .st-table td, .st-table th {   padding: 8px;
```

```
heUrl_ca2b <- "https://surfdrive.surf.nl/files/index.php/s/DOGvC9BFm945QF1/download"
airbnb <- read_dta ("edelman2017_v1112.dta")
```

1. Define the two potential outcomes for a particular guest, Y(0,i) and Y(1,i). Provide one or more reasons why someone's last name may affect the chance of being accepted by a host.

Y(0,i) = not being accepteded to be a guest Y(1,i) = being accepted to be a guest

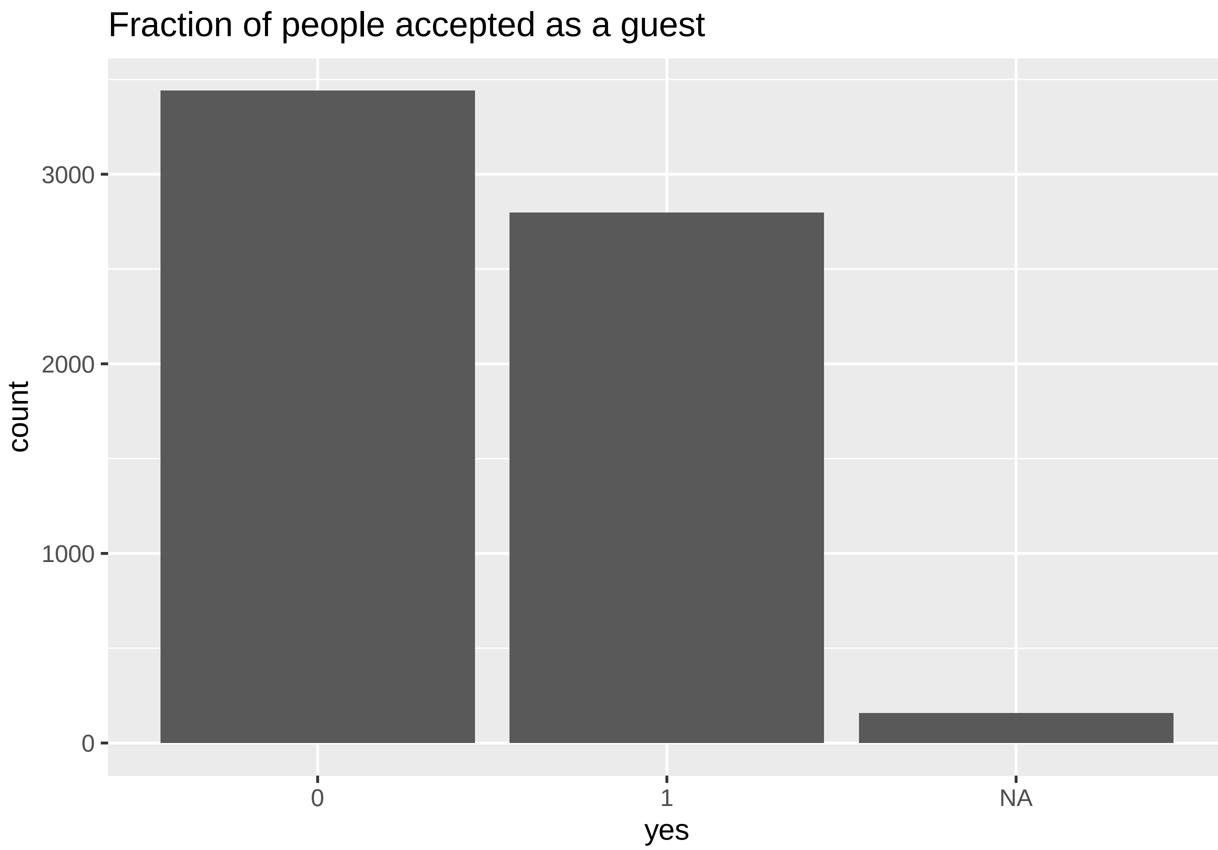
Someone's last name is passed down through generations and so can have connections with a persons culture/history/race. Humans attach charateristics to groups of people, with some attaching negative prejudices that would influence the decision to accept the guest before actually meeting the peopl. This can usually be explained through selection bias, where a person opinions about a particular few people and applies it to be true for all cases in the population. The media/institutions in power also help connect qualitities of a few to the whole population in a way to influence and control peoples opinions.

2. Browse the data set. Make sure to browse ALL variables, by selecting a further set of columns at the top. Then provide:

(a) a histogram of the outcome variable;

```
ggplot(airbnb, aes(x=as.factor(yes)))+
geom_histogram(stat="count")+
labs(x='yes', y='count', title='Fraction of people accepted as a guest')
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



(b) the overall mean acceptance rate.

```
summary(airbnb$yes)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.0000  0.0000  0.0000  0.4484  1.0000  1.0000     157
```

3. Check for balance on two covariates: whether the host is African-American and whether the host is male (see Table 1 above) - by way of cross tabs. Test for a difference between treatment and control for whether the host is male (only).

```
ctable(airbnb$guest_black, airbnb$host_race_black)
```

```
## ### Cross-Tabulation, Row Proportions
## #### guest_black * host_race_black
## **Data Frame:** airbnb
##
## |             |                 |              |           |              |
## |------------:|----------------:|-------------:|----------:|-------------:|
## |             | host_race_black |            0 |         1 |        Total |
## | guest_black |                 |              |           |              |
## |           0 |                 | 2952 ( 92.2%) | 251 (7.8%) | 3203 (100.0%) |
## |           1 |                 | 2938 ( 92.2%) | 249 (7.8%) | 3187 (100.0%) |
## |      \<NA\> |                 |    2 (100.0%) |   0 (0.0%) |    2 (100.0%) |
## |       Total |                 | 5892 ( 92.2%) | 500 (7.8%) | 6392 (100.0%) |
```

```
ctable(airbnb$guest_black, airbnb$host_gender_M)
```

```
## ### Cross-Tabulation, Row Proportions
## #### guest_black * host_gender_M
```

```
## **Data Frame:** airbnb
##
## |            |              |              |             |              |
## |-----------:|-------------:|-------------:|------------:|-------------:|
## |            | host_gender_M |            0 |           1 |        Total |
## | guest_black |              |              |             |              |
## |          0 |              | 2250 ( 70.2%) |  953 (29.8%) | 3203 (100.0%) |
## |          1 |              | 2234 ( 70.1%) |  953 (29.9%) | 3187 (100.0%) |
## |     \<NA\> |              |    2 (100.0%) |    0 ( 0.0%) |    2 (100.0%) |
## |      Total |              | 4486 ( 70.2%) | 1906 (29.8%) | 6392 (100.0%) |
```

Test for a difference between treatment and control for whether the host is male (only).

```r
t.test(airbnb$host_gender_M~airbnb$guest_black)
```

```
##
##  Welch Two Sample t-test
##
## data:  airbnb$host_gender_M by airbnb$guest_black
## t = -0.13048, df = 6387.7, p-value = 0.8962
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02393636  0.02094889
## sample estimates:
## mean in group 0 mean in group 1
##       0.2975336       0.2990273
```

There was a statistically significant difference found at a significance level of 5%.

4. Minimum sample size. Let us put ourselves in the position of the authors before the data were collected. How many guests would they have to include in their sample at the minimum to be able to detect a treatment effect, assuming that there is a treatment effect? Assume that they want to have statistical power of 80%.

```r
airbnb %>%
  filter(guest_black==0) %>%
  summarise(mean=mean(yes, na.rm=TRUE), sd=sd(yes, na.rm=TRUE))
```

mean	sd
0.4877894	0.4999302
Say that a 5	percentage-points change in the acceptance rate is the minimum effect of this treatment.
Effect Size:	$0.05/0.5 = 0.1$ - effect of $d = 0.1$ standard deviations is considered small.

```r
pwr.t.test(n = NULL, d = 0.1, sig.level = 0.05, power = 0.8,
           type = c("two.sample"), alternative="two.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 1570.733
##               d = 0.1
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
```

```
## NOTE: n is number in *each* group
```

The minimum sample size is n = 1571

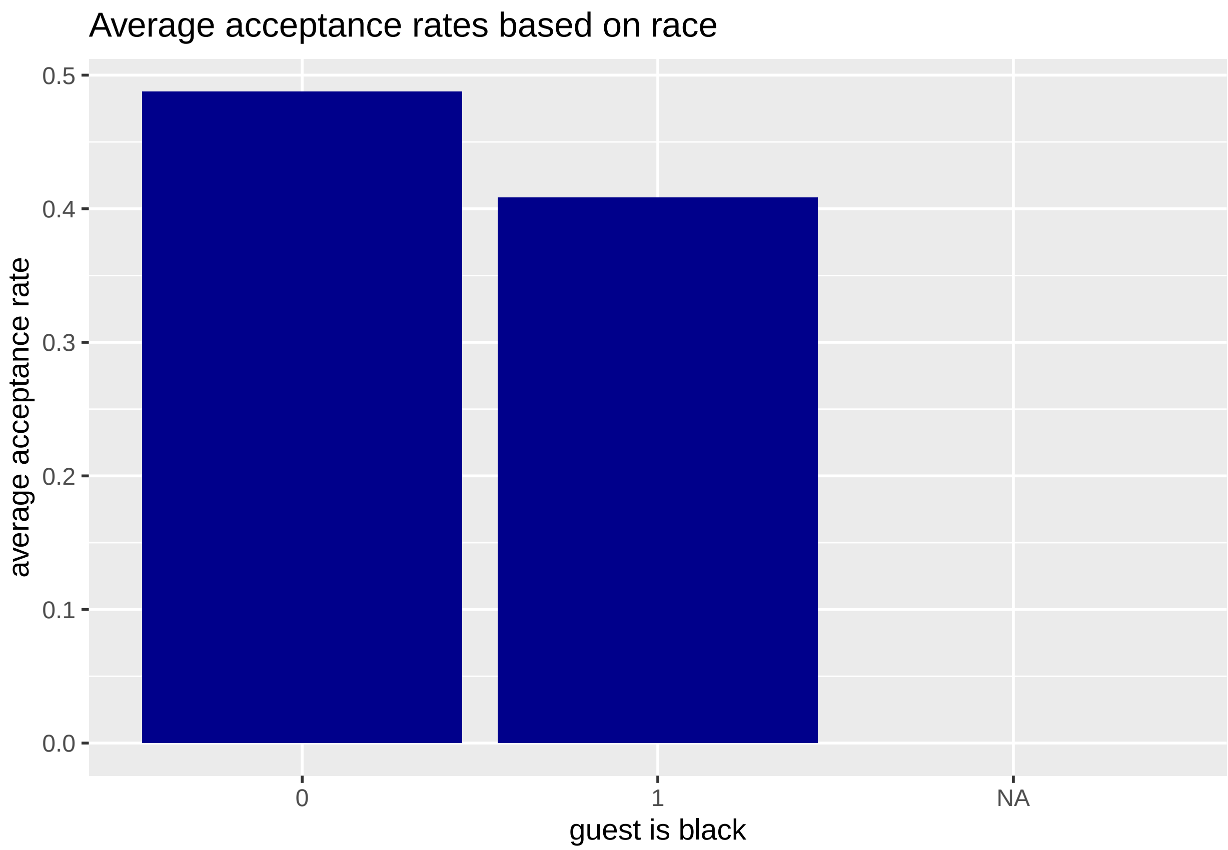
5. Graphically explore the treatment effect. Create a bar graph showing the difference in the outcome variable by the treatment dummy.

```
airbnb_peek <- airbnb %>% group_by(guest_black) %>%
  summarise(yes_mean=mean(yes, na.rm=TRUE))
airbnb_peek
```

guest_black	yes_mean
0	0.4877894
1	0.4081765
NA	NaN

```
ggplot(airbnb_peek, aes(y=yes_mean, x=as.factor(guest_black))) + geom_bar(stat='identity', fill='darkblu
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



For African-American/Black people, there is a lower average number of acceptances as a guest.

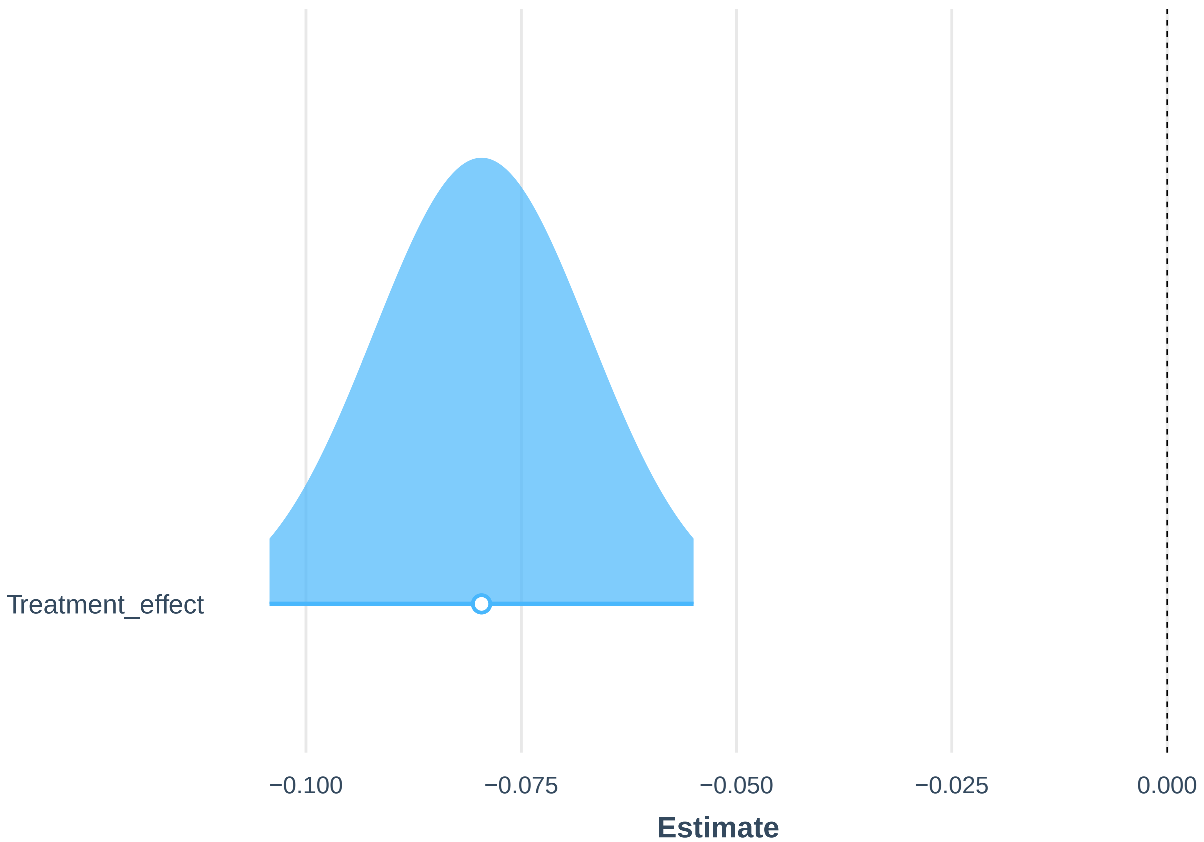
6. Estimate the treatment effect. Replicate the three columns of Edelman and Luca 2017, Table 2, p. 8.

```
#Regression 1 with only guest_black included
reg1 <- lm(yes ~ guest_black, data=airbnb)
summ(reg1, confint=TRUE)
```

```
## MODEL INFO:
```

```
## Observations: 6235 (157 missing obs. deleted)
## Dependent Variable: yes
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,6233) = 40.18, p = 0.00
## R² = 0.01
## Adj. R² = 0.01
##
## Standard errors: OLS
## ---------------------------------------------------------
##                         Est.    2.5%   97.5%   t val.      p
## ----------------- -------- -------- -------- --------- ------
## (Intercept)            0.49    0.47    0.51    55.24   0.00
## guest_black           -0.08   -0.10   -0.05    -6.34   0.00
## ---------------------------------------------------------
```

```
treatment_effect <- c("Treatment_effect"="guest_black")
plot_summs(reg1, scale = TRUE, coefs = treatment_effect, plot.distributions = TRUE)
```




```
#Regression 2 with guest is black, host is black, host is male included
reg2 <- lm(yes ~ guest_black + host_race_black + host_gender_M, data=airbnb)
summ(reg2, confint=FALSE)
```

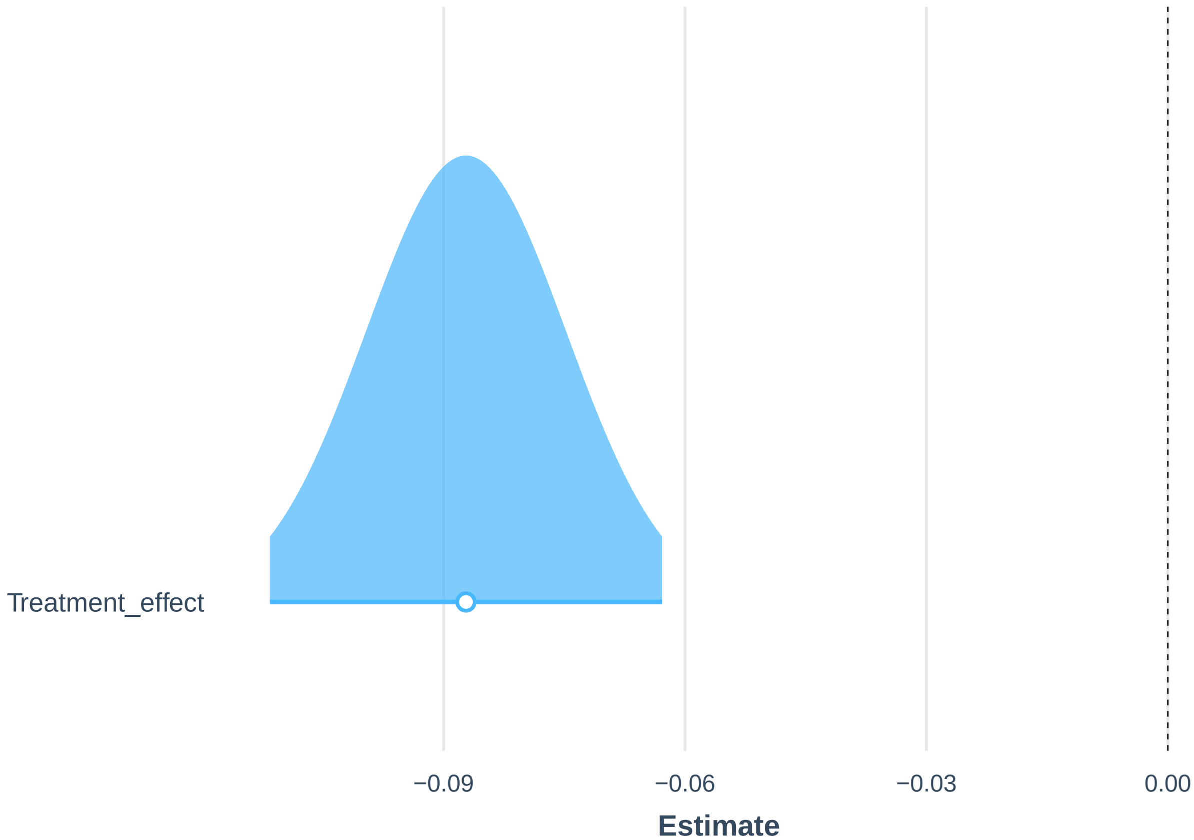
```
## MODEL INFO:
## Observations: 6235 (157 missing obs. deleted)
## Dependent Variable: yes
## Type: OLS linear regression
```

```
##
## MODEL FIT:
## F(3,6231) = 20.50, p = 0.00
## R² = 0.01
## Adj. R² = 0.01
##
## Standard errors: OLS
## ----------------------------------------------------
##                          Est.   S.E.   t val.      p
## --------------------- -------- ------ -------- ------
## (Intercept)              0.50   0.01    50.49   0.00
## guest_black             -0.08   0.01    -6.35   0.00
## host_race_black          0.07   0.02     2.94   0.00
## host_gender_M           -0.05   0.01    -3.67   0.00
## ----------------------------------------------------
```

```r
#Regression 3 with guest is black, host is black, host is male included, host has multiple listings, sh
reg3 <- lm(yes ~ guest_black + host_race_black + host_gender_M + multiple_listings +
               shared_property + ten_reviews + log_price, data=airbnb)
summ(reg3, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 6168 (224 missing obs. deleted)
## Dependent Variable: yes
## Type: OLS linear regression
##
## MODEL FIT:
## F(7,6160) = 36.68, p = 0.00
## R² = 0.04
## Adj. R² = 0.04
##
## Standard errors: OLS
## ------------------------------------------------------------
##                            Est.    2.5%   97.5%   t val.      p
## ----------------------- -------- -------- -------- --------- ------
## (Intercept)                0.76    0.64    0.87    12.92   0.00
## guest_black               -0.09   -0.11   -0.06    -7.02   0.00
## host_race_black            0.09    0.05    0.14     3.96   0.00
## host_gender_M             -0.05   -0.07   -0.02    -3.52   0.00
## multiple_listings          0.06    0.04    0.09     4.46   0.00
## shared_property           -0.07   -0.10   -0.04    -4.43   0.00
## ten_reviews                0.12    0.09    0.15     9.04   0.00
## log_price                 -0.06   -0.08   -0.04    -5.74   0.00
## ------------------------------------------------------------
```

```r
treatment_effect <- c("Treatment_effect"="guest_black")
plot_summs(reg3, scale = TRUE, coefs = treatment_effect, plot.distributions = TRUE)
```



The regression analysis indicates a lower acceptance rate for those who have a Black/African-American name as opposed to those who have not. The first two regressions report an estimated change of -8 percentage points in acceptances for those who have a Black/African-American name.