CA2bVdVeen

Ties van der Veen

7-9-2019

Rebecca Costello - Ties van der Veen - Niels van Opstal

I. Introduction to the assignment

```
library(foreign)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggdag)
```

```
##
## Attaching package: 'ggdag'
```

```
## The following object is masked from 'package:ggplot2':
##
##     expand_scale
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(dplyr)
library(tinytex)
library(jtools)
library(huxtable)
```

```
##
## Attaching package: 'huxtable'
```

```
## The following objects are masked from 'package:ggdag':
##
##     label, label<-
```

```
## The following object is masked from 'package:dplyr':
##
##     add_rownames

## The following object is masked from 'package:purrr':
##
##     every

## The following object is masked from 'package:ggplot2':
##
##     theme_grey
```

```
library(summarytools)
```

```
## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp

##
## Attaching package: 'summarytools'

## The following objects are masked from 'package:huxtable':
##
##     label, label<-

## The following objects are masked from 'package:ggdag':
##
##     label, label<-

## The following object is masked from 'package:tibble':
##
##     view
```

```
library(ggstance)
```

```
##
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh
```

```
library(pwr)
library(knitr)
library(lemon)
```

```
##
## Attaching package: 'lemon'

## The following object is masked from 'package:purrr':
##
##     %||%
```

```
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;
```

```
theUrl_ca2b <- "https://surfdrive.surf.nl/files/index.php/s/DOGvC9BFm945QF1/download"

airbnb <- read.dta (file = theUrl_ca2b)
```

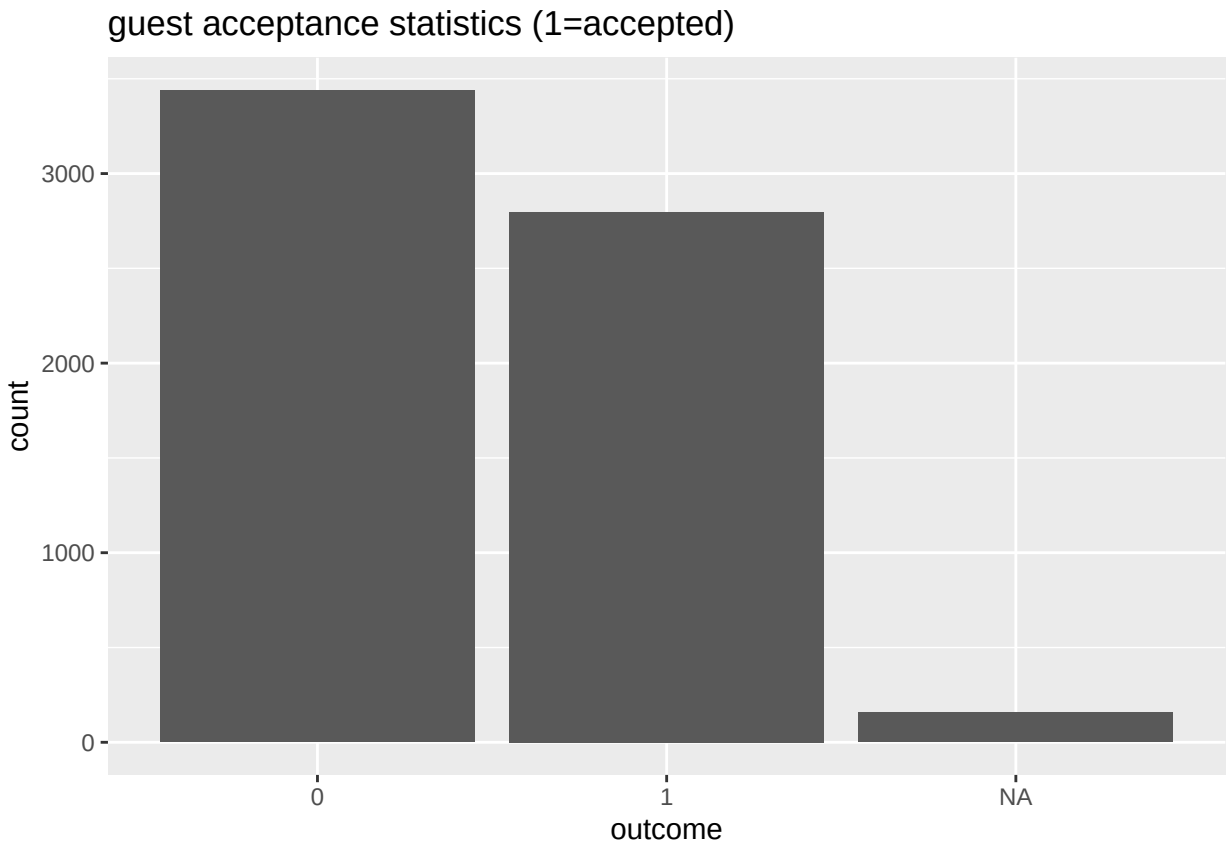
II. What to submit

(1) Y(0,i) = guest with a non-African American name Y(1,i) = guest with a distinctly African American name The name of a guest may influence the chance of being accepted by a host through influences of hearsay, negative experiences, slight racism, or just safety in known names.

(2) (a)

```
ggplot(airbnb, aes(x=as.factor(yes)))+
  geom_histogram(stat='count')+
  labs(x='outcome', y='count', title='guest acceptance statistics (1=accepted)')
```

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



(b)

```
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)

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

```
## ### Cross-Tabulation, Row Proportions
## #### yes * guest_black
## **Data Frame:** airbnb
##
## |        |             |              |              |          |               |
## |-------:|------------:|-------------:|-------------:|---------:|--------------:|
## |        | guest_black |            0 |            1 |   \<NA\> |         Total |
## |    yes |             |              |              |          |               |
## |      0 |             | 1615 (47.0%) | 1824 (53.0%) | 0 (0.0%) | 3439 (100.0%) |
## |      1 |             | 1538 (55.0%) | 1258 (45.0%) | 0 (0.0%) | 2796 (100.0%) |
## | \<NA\> |             |   50 (31.8%) |  105 (66.9%) | 2 (1.3%) |  157 (100.0%) |
## |  Total |             | 3203 (50.1%) | 3187 (49.9%) | 2 (0.0%) | 6392 (100.0%) |
```

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

```
## ### Cross-Tabulation, Row Proportions
## #### yes * host_gender_M
## **Data Frame:** airbnb
##
## |        |               |             |             |               |
## |-------:|--------------:|------------:|------------:|--------------:|
## |        | host_gender_M |           0 |           1 |         Total |
## |    yes |               |             |             |               |
## |      0 |               | 2353 (68.4%) | 1086 (31.6%) | 3439 (100.0%) |
## |      1 |               | 2028 (72.5%) |  768 (27.5%) | 2796 (100.0%) |
## | \<NA\> |               |  105 (66.9%) |   52 (33.1%) |  157 (100.0%) |
## |  Total |               | 4486 (70.2%) | 1906 (29.8%) | 6392 (100.0%) |
```

```
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
```

(4)

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

mean	sd
0.488	0.5

Calculate d by taking the percentage-point change of 5% that we assume to be the minimum effect, and divide this by the standard deviation of the outcome: $0.05/0.5 = 0.1$. Thus our d will be 0.1.

```
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
```

Thus the minimum sample size (n) is 1571 (per group)

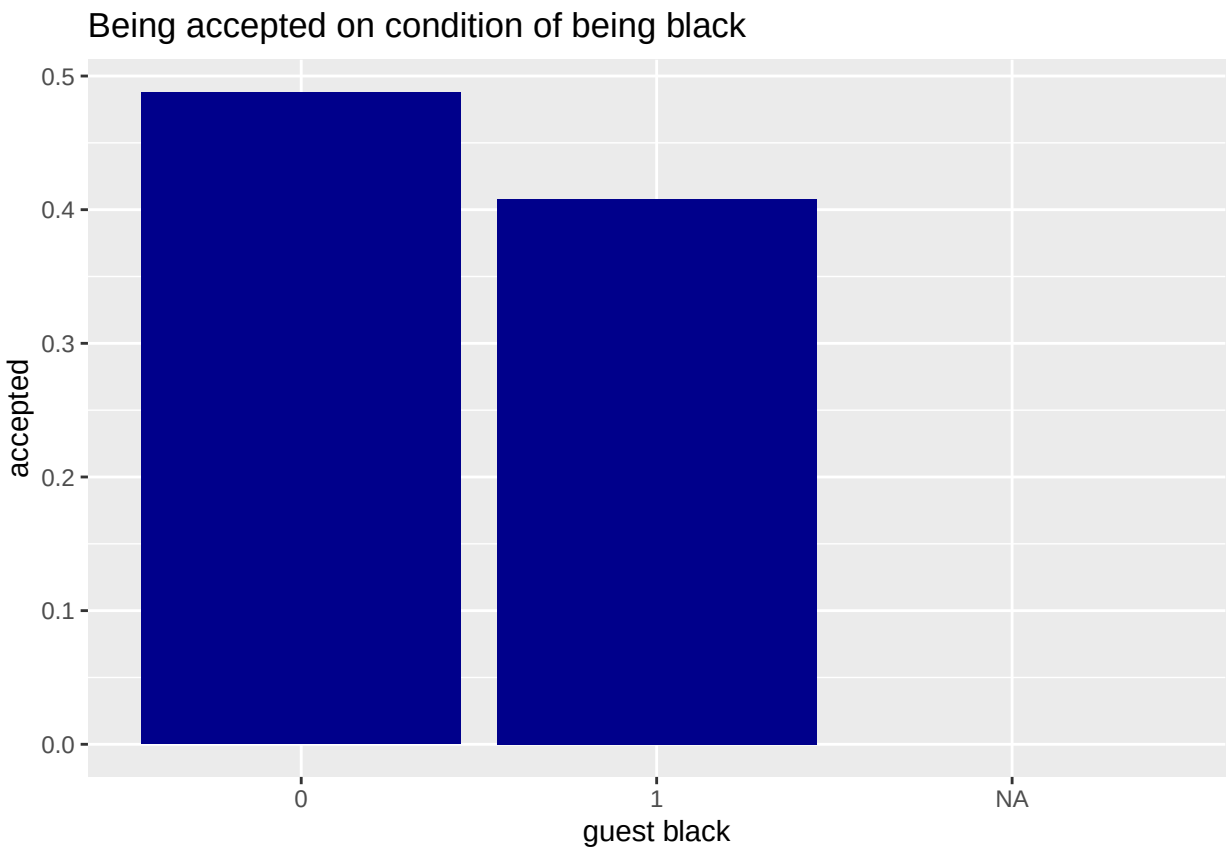
(5)

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

guest_black	yes_mean
0	0.488
1	0.408
	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).
```



(6)

```
reg2b1 <- lm(yes ~ guest_black, data=airbnb)
summ(reg2b1, 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
## ------------------------------------------------------------
```

```
reg2b2 <- lm(yes ~ guest_black + host_race_black + host_gender_M, data=airbnb)
summ(reg2b2, 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
## -----------------------------------------------------
```

```
reg2b3 <- lm(yes ~ guest_black + host_race_black + host_gender_M + multiple_listings +
               shared_property + ten_reviews + log_price, data=airbnb)
summ(reg2b3, 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
## ------------------------------------------------------------------
```

The regression results suggest a lower acceptance rate for those who are african american as compared to those who are not. This result does seem to be small on its own however, with an estimate of -0.09. Comparatively, it is bigger than the 5% minimum effect rate that we assumed earlier, so the result is large when related to that number.