CA2a

Ties van der Veen

6-9-2019

Rebecca Costello - Ties van der Veen - Niels van Opstal

I.Introduction to the assignment

Packages

```
options(repos="https://cran.rstudio.com")
install.packages("jtools")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'jtools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
install.packages("huxtable")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'huxtable' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
install.packages("ggstance")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'ggstance' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
install.packages("summarytools")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'summarytools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
install.packages("pwr")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)

## package 'pwr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
install.packages("knitr")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)

## package 'knitr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
install.packages("lemon")
```

```
## Installing package into 'C:/Users/tiess/OneDrive/Documenten/R/win-library/3.6'
## (as 'lib' is unspecified)

## package 'lemon' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpiuf7pP\downloaded_packages
```

```
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()
```

```r
library(ggdag)
```

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

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

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

```r
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
```

```r
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;
```

Data

```
theUrl_ca2a <- "https://surfdrive.surf.nl/files/index.php/s/ULZJ0bBbphCttpG/download"

students <- read.dta (file = theUrl_ca2a)
```

II. Potential outcomes

(a) Y(0,i) = asked for risk perception, then recall the last instance of bicycle theft Y(1,i) = asked to recall last instance of bicycle theft, then risk perception
(b) Their perception of bicycle theft could be affected because they were reminded of the last time this happened. This makes the memory more salient, which makes them consider the risk more readily.

III. Descriptive statistics

```
summary(students)
```

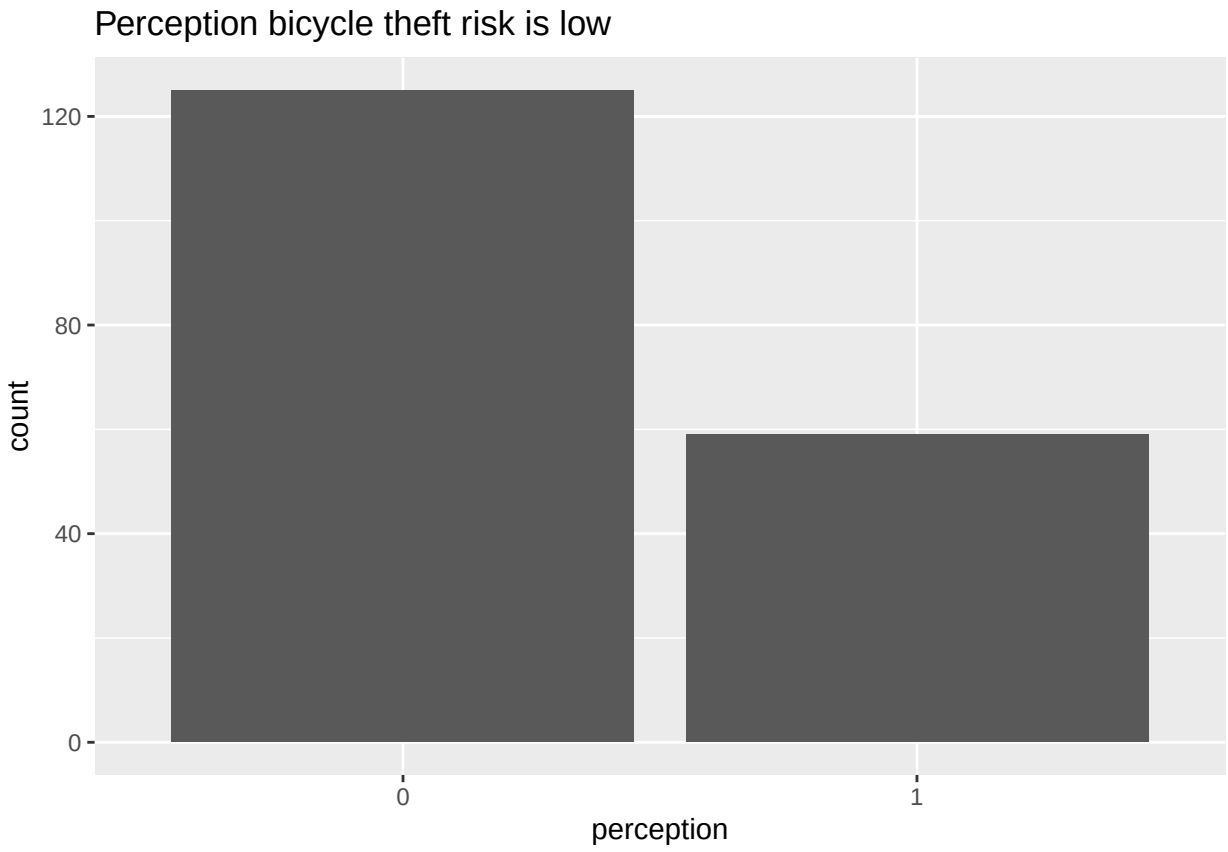
```
##   frequentuser    bicyclestolen_ever     female       international
##  Min.   :0.0000   Min.   :0.0000     Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000     Median :0.0000   Median :0.0000
##  Mean   :0.8913   Mean   :0.3315     Mean   :0.3913   Mean   :0.4891
##  3rd Qu.:1.0000   3rd Qu.:1.0000     3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000     Max.   :1.0000   Max.   :1.0000
##  moved_notrecent    treatment     perception_person_low     age20
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000        Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000        1st Qu.:0.00000
##  Median :0.0000   Median :1.000   Median :0.0000        Median :0.00000
##  Mean   :0.2663   Mean   :0.538   Mean   :0.3207        Mean   :0.03261
##  3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000        3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000        Max.   :1.00000
##    cohort2019
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3533
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
summary(students$perception_person_low)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.3207  1.0000  1.0000
```

```
ggplot(students, aes(x=as.factor(perception_person_low)))+
  geom_histogram(stat="count")+
  labs(x='perception', y='count', title='Perception bicycle theft risk is low')
```

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




```
summary(is.na(students$perception_person_low))
```

```
##    Mode   FALSE
## logical     184
```

IV. Balance check

```
ctable(students$treatment, students$female)
```

```
## ### Cross-Tabulation, Row Proportions
## #### treatment * female
## **Data Frame:** students
##
## |           |        |             |            |              |
## |----------:|-------:|------------:|-----------:|-------------:|
## |           | female |           0 |          1 |        Total |
## | treatment |        |             |            |              |
## |         0 |        |  49 (57.6%) | 36 (42.4%) |  85 (100.0%) |
## |         1 |        |  63 (63.6%) | 36 (36.4%) |  99 (100.0%) |
## |     Total |        | 112 (60.9%) | 72 (39.1%) | 184 (100.0%) |
```

```r
ctable(students$treatment, students$international)
```

```
## ### Cross-Tabulation, Row Proportions
## #### treatment * international
## **Data Frame:** students
##
## |           |               |           |           |             |
## |----------:|--------------:|----------:|----------:|------------:|
## |           | international |         0 |         1 |       Total |
## | treatment |               |           |           |             |
## |         0 |               | 42 (49.4%) | 43 (50.6%) |  85 (100.0%) |
## |         1 |               | 52 (52.5%) | 47 (47.5%) |  99 (100.0%) |
## |     Total |               | 94 (51.1%) | 90 (48.9%) | 184 (100.0%) |
```

```r
ctable(students$treatment, students$moved_notrecent)
```

```
## ### Cross-Tabulation, Row Proportions
## #### treatment * moved_notrecent
## **Data Frame:** students
##
## |           |                |            |           |             |
## |----------:|---------------:|-----------:|----------:|------------:|
## |           | moved_notrecent |          0 |         1 |       Total |
## | treatment |                |            |           |             |
## |         0 |                |  62 (72.9%) | 23 (27.1%) |  85 (100.0%) |
## |         1 |                |  73 (73.7%) | 26 (26.3%) |  99 (100.0%) |
## |     Total |                | 135 (73.4%) | 49 (26.6%) | 184 (100.0%) |
```

```r
ctable(students$treatment, students$age20)
```

```
## ### Cross-Tabulation, Row Proportions
## #### treatment * age20
## **Data Frame:** students
##
## |           |       |            |         |             |
## |----------:|------:|-----------:|--------:|------------:|
## |           | age20 |          0 |       1 |       Total |
## | treatment |       |            |         |             |
## |         0 |       |  82 (96.5%) | 3 (3.5%) |  85 (100.0%) |
## |         1 |       |  96 (97.0%) | 3 (3.0%) |  99 (100.0%) |
## |     Total |       | 178 (96.7%) | 6 (3.3%) | 184 (100.0%) |
```

```r
t.test(students$female~students$treatment)
```

```
##
##  Welch Two Sample t-test
##
## data:  students$female by students$treatment
## t = 0.8252, df = 176.23, p-value = 0.4104
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.0833447  0.2031308
## sample estimates:
## mean in group 0 mean in group 1
##       0.4235294       0.3636364
```

VI. Statistical power

```
students %>%
  filter(treatment==0) %>%
  summarise(mean=mean(perception_person_low), sd=sd(perception_person_low))
```

mean	sd
0.4	0.493

```
pwr.t.test(n = NULL, d = 0.2, sig.level = 0.05, power = 0.8, type = c("two.sample"), alternative="two.si
```

```
##
##      Two-sample t test power calculation
##
##               n = 393.4057
##               d = 0.2
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
pwr.t.test(n = 92, d = 0.2, sig.level = 0.05, power = NULL, type = c("two.sample"), alternative="two.sid
```

```
##
##      Two-sample t test power calculation
##
##               n = 92
##               d = 0.2
##       sig.level = 0.05
##           power = 0.2711829
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

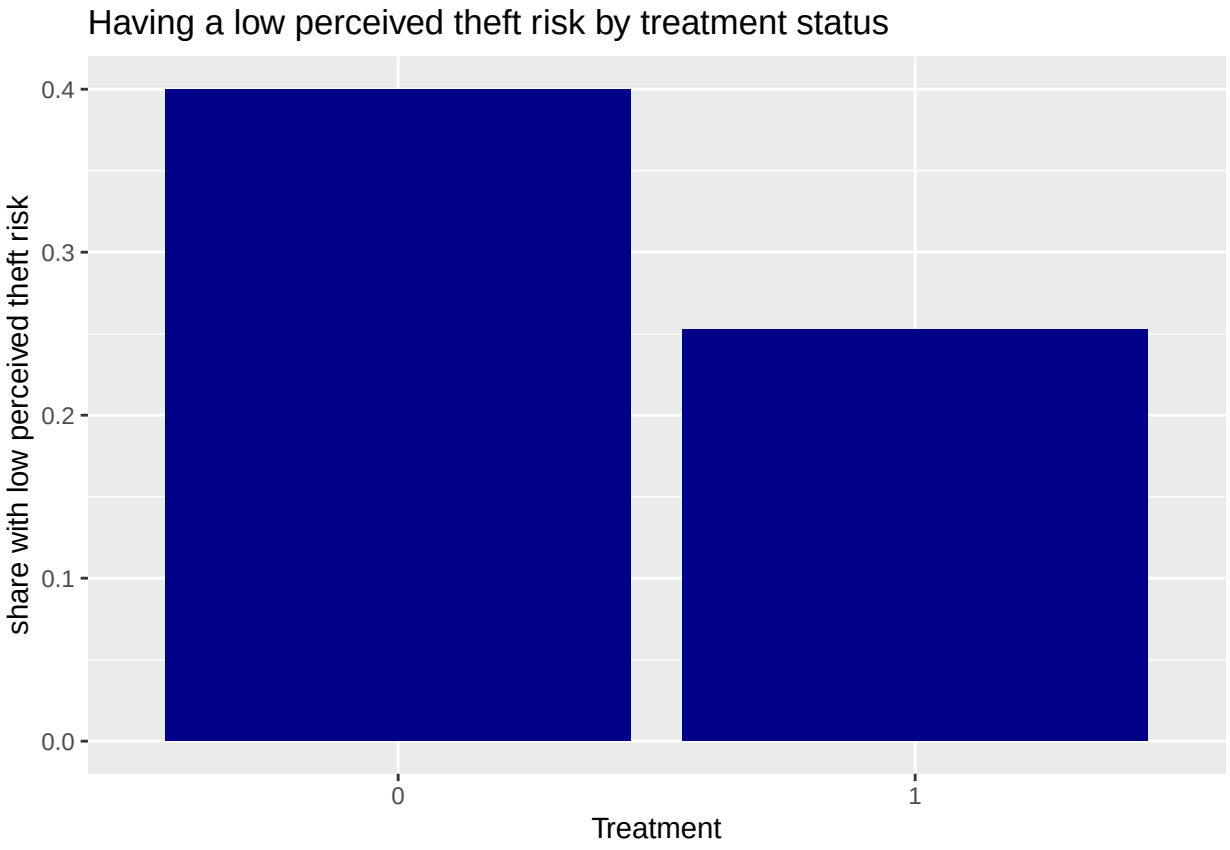
VI. Estimating treatment effect in a randomized trial

(a)

```r
students_peek <- students %>% group_by(treatment) %>%
  summarise(perception_person_low_mean=mean(perception_person_low))
students_peek
```

treatment	perception_person_low_mean
0	0.4
1	0.253

```r
ggplot(students_peek, aes(y=perception_person_low_mean, x=as.factor(treatment))) + geom_bar(stat='ident
```



The graph suggests that the control group has more people with a low perception of theft risk.

(b) Y = a + B*treatment + e

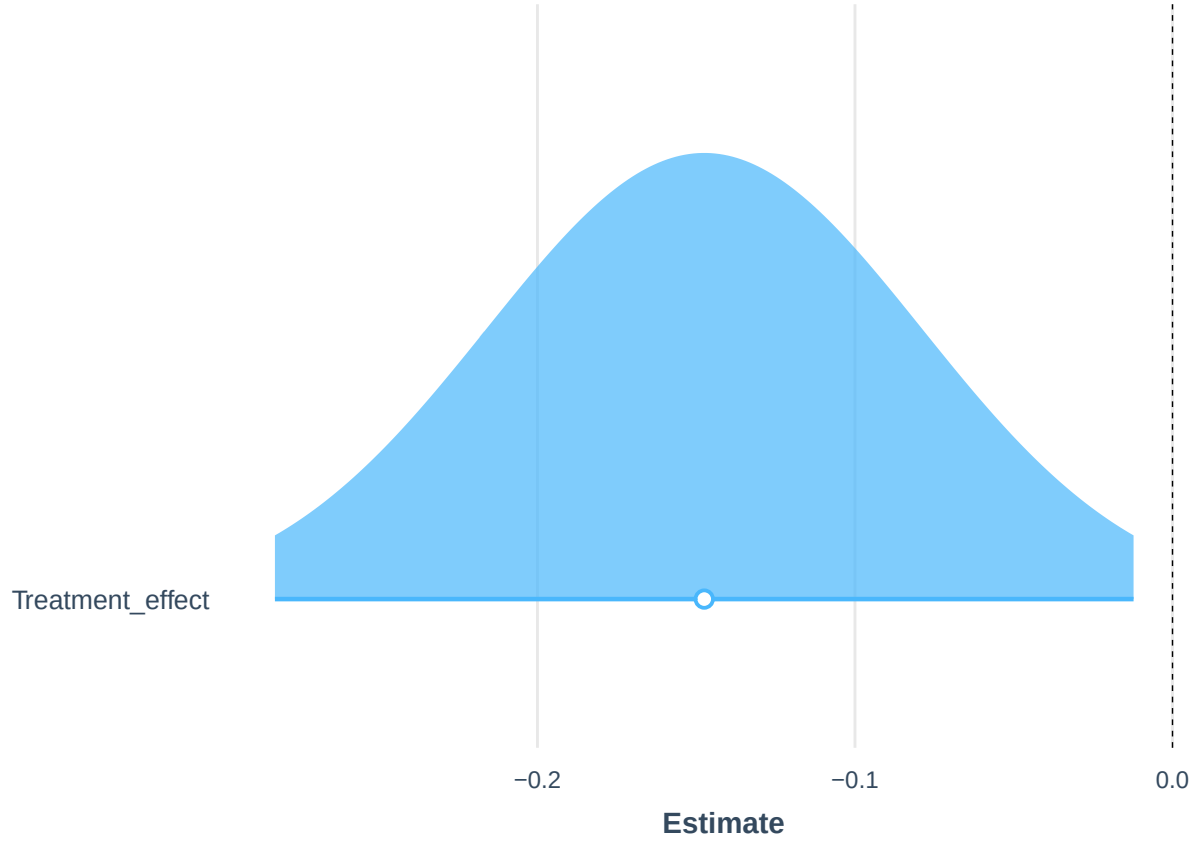
(c)

```r
reg1 <- lm(perception_person_low ~ treatment, data=students)
summ(reg1, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 184
## Dependent Variable: perception_person_low
```

```
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,182) = 4.63, p = 0.03
## R² = 0.02
## Adj. R² = 0.02
##
## Standard errors: OLS
## ------------------------------------------------------------
##                       Est.    2.5%   97.5%   t val.      p
## ----------------- ------- ------- ------- -------- ------
## (Intercept)           0.40    0.30    0.50     7.96   0.00
## treatment            -0.15   -0.28   -0.01    -2.15   0.03
## ------------------------------------------------------------
```

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



(d) Yes. The regression shows that p = 0.03, so we can reject the null hypothesis at 5% confidence level.

(e) The regression results suggest a lower perception of bicycle theft for those who were reminded of bicycle theft as compared to those who were not.

(f) One can check this the following way: (Estimated treatment effect) / (baseline mean of outcome variable) * 100%. In this case: -0.15/0.2525253*100% = -59.4%. Thus, the treatment effect seems to be rather large.

(g) R^2 in the regression is 0.02, so the size seems to be appropriate

(h) No, because the random selection into control/treatment accounted for this (control also has people who have never had their bike stolen).