Problem Set #1 - DATASCI W241

Greg Ceccarelli

Sept 15, 2015

1. On the notation of potential outcomes

a. Concisely, $Y_i(1)$ is the outcome if the subject i is exposed to the treatment
b. $E[Y_i(1)|d_i = 0]$ denotes the expectation of $Y_i(1)$ (i.e. the outcome of the subject i exposed to the treatment) when one subject is selected at random from the subjects that weren't treated (i.e. the counterfactual). In practice we won't actually observe $Y_i(1)$ for an untreated subject.
c. $E[Y_i(1)|d_i = 1]$ is the expectation of $Y_i(1)$ (i.e. the outcome) when one subject is selected at random from the subjects that were treated whereas $E[Y_i(1)]$ is the expected $Y_i(1)$ potential outcome for the entire set of subjects.
d. (extra credit)

FE, exercise 2.2

```
#set up working envrionment
setwd('/Users/ceccarelli/MIDS/DATASCI_W241/Async Material and Sample Files/Chapter 2/')
#clear variables
rm( list = ls() )
#read in tabular data
potential_outcomes.data <- read.csv("GerberGreenBook_Chapter2_Table_2_1.csv", sep=",", header = TRUE)
#create shorthand reference
p<-potential_outcomes.data
##View(potential_outcomes.data)
#Define E[Yi(1)]
eYi<-sum(p$Yi.1.)/nrow(p)
#Define E[Yi(0)]
eYo<-sum(p$Yi.0.)/nrow(p)
#Define  E[Yi(0) - Yi(1)]
ate<-sum(p$Yi.0-p$Yi.1.)/nrow(p)
#Test equivalence between E[Yi(0)] - E[Yi(1)] = E[Yi(0) - Yi(1)]
(eYo - eYi) == ate
```

```
## [1] TRUE
```

FE, exercise 2.3

a.

```
#set up working envrionment
p.subset <- p[c(2:3)]
#create matrix to count observations by variable pairs
mat<-table(p.subset)
#print matrix
mat
```

```
##      Yi.1.
## Yi.0. 15 20 30
##    10  1  1  0
##    15  2  0  1
##    20  1  0  1
```

b.

```
#count number of obs in matrix
num_obs = sum(mat[,1]+mat[,2]+mat[,3])
#calc percentage of subjects in each cell (joint freq distribution)
mat.jfd <- mat / num_obs
#print new matrix
mat.jfd
```

```
##      Yi.1.
## Yi.0.        15        20        30
##    10 0.1428571 0.1428571 0.0000000
##    15 0.2857143 0.0000000 0.1428571
##    20 0.1428571 0.0000000 0.1428571
```

c.

```
#marginal distribution of Yi(1)
mat.jfd[1,]+mat.jfd[2,]+mat.jfd[3,]
```

```
##        15        20        30
## 0.5714286 0.1428571 0.2857143
```

```
#add as row in matrix
mat.jfd <- rbind(mat.jfd, mat.jfd[1,]+mat.jfd[2,]+mat.jfd[3,])
#Update Rownames
rownames(mat.jfd)[4]<-"Yi.1"
mat.jfd
```

```
##             15        20        30
## 10   0.1428571 0.1428571 0.0000000
## 15   0.2857143 0.0000000 0.1428571
## 20   0.1428571 0.0000000 0.1428571
## Yi.1 0.5714286 0.1428571 0.2857143
```

d.

```
#marginal distribution of Yi(0)
mat.jfd[,1]+mat.jfd[,2]+mat.jfd[,3]
```

```
##        10        15        20      Yi.1
## 0.2857143 0.4285714 0.2857143 1.0000000
```

```r
#add as column in matrix
mat.jfd <- cbind(mat.jfd, mat.jfd[,1]+mat.jfd[,2]+mat.jfd[,3])
#update colnames
colnames(mat.jfd)[4]<-"Yi.0"
mat.jfd
```

```
##             15        20        30      Yi.0
## 10   0.1428571 0.1428571 0.0000000 0.2857143
## 15   0.2857143 0.0000000 0.1428571 0.4285714
## 20   0.1428571 0.0000000 0.1428571 0.2857143
## Yi.1 0.5714286 0.1428571 0.2857143 1.0000000
```

e.

```r
#use table to calculate condition expectation that E[Yi(0)|Yi(1)>15]
#
```

f.

```r
#use table to calculate condition expectation that E[Yi(1)|Yi(0)>15]
#
```

4. More practice with potential outcomes

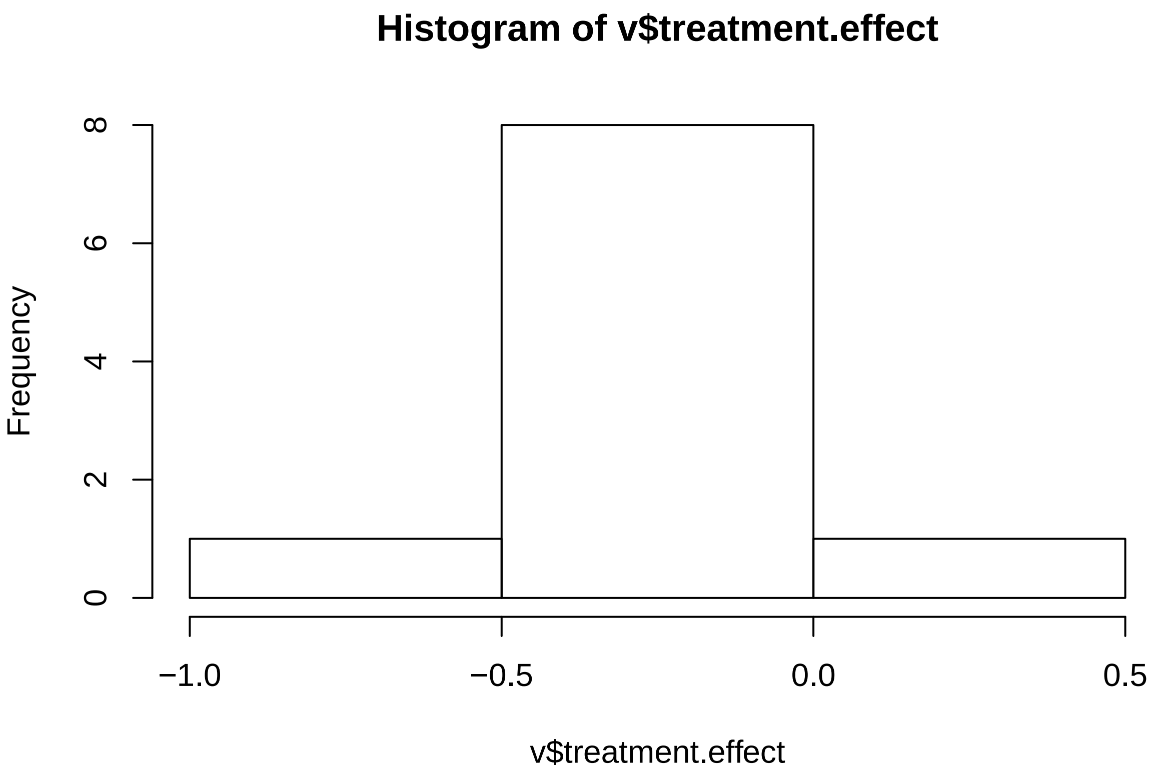
a.

```r
setwd('/Users/ceccarelli/MIDS/DATASCI_W241/Async Material and Sample Files/Chapter 2/')
#read in visual acuity data
visualacuity.data <- read.csv("VisualAcuity.csv", sep=",", header = TRUE)
#create shorthand reference
v<-visualacuity.data
#compute treatment effect
v$treatment.effect <- v$Yi.1.-v$Yi.0.
v$treatment.effect
```

```
##  [1]  0.0  0.5  0.0  0.0 -0.9  0.0  0.0  0.0  0.0  0.0
```

b.

```r
#quickly review distribution of treatment effects
hist(v$treatment.effect)
```



Paragraph to add

c.

```
#test computing true average treatment effect
sum(v$treatment.effect) / nrow(v) == sum(v$Yi.1.)/nrow(v)-sum(v$Yi.0.)/nrow(v)
```

```
## [1] FALSE
```

```
#print ATE
sum(v$treatment.effect) / nrow(v)
```

```
## [1] -0.04
```

d.

```
#generate indexes
rows<-nrow(v)
even_indexes<-seq(2,rows,2)
odd_indexes<-seq(1,rows,2)
#define new dataset
v.odd <- v[1:3]
#assign odd children to treatment by setting their hypothetical control value to NA
v.odd[odd_indexes,2]<-NA
#assign even children to control by setting their hypothetical treatment value to NA
v.odd[even_indexes,3]<-NA
#Estimated average based on observed data
v.odd.ate <-  mean(v.odd$Yi.1.,na.rm = TRUE) -  mean(v.odd$Yi.0.,na.rm = TRUE)
#print ate
v.odd.ate
```

```
## [1] -0.06
```

e. The ATE -.06 of this new experiment differs by .02 when compared to ATE -.04 of the hypothetical example. Intuitively, a "random" allocation of children to treatment effects will produce groups of children that have different average potential outcomes.

f.

g.

h.

FE, exercise 2.5

a. The problem is how to randomize treatment for six subjects asked to donate time to an adult literacy program.

b.

c.

FE, exercise 2.6

FE, exercise 2.8