36-402 Homework 3

James "Morgan" Hawkins

Friday Feb 10, 2023

Problem 1

Problem 1 (c)

```
rx <- function(b){
  if(b == 0){
    return(runif(1,0,4))
  }else{
    return(rnorm(1,-2,2))
  }

}

set.seed(99)
sample.z <- rbinom(10000,1,.5)
sample.x <-sapply(sample.z,rx)

mean(sample.x)
```

```
## [1] 0.01001486
```

```
sd(sample.x)^2
```

```
## [1] 6.674159
```

We expect our sample to have a mean of 0 and a variance of 6.666. After drawing 10,000 smaples from the distribution of x we get a sample mean of .0100 and a sample variance of 6.6742. This reinfoces our answers for part a and b.

Problem 4

Problem 4 (a)

```
1198/(1493+1198)
```

```
## [1] 0.4451877
```

```r
557/(557+1278)
```

```
## [1] 0.3035422
```

44.52% of male applicants were accepted to Berkeley while only 30.35% of female applicants were accepted.

Problem 4 (b)

```r
ucb = data.frame(UCBAdmissions)
ucb
```

```
##       Admit Gender Dept Freq
## 1  Admitted   Male    A  512
## 2  Rejected   Male    A  313
## 3  Admitted Female    A   89
## 4  Rejected Female    A   19
## 5  Admitted   Male    B  353
## 6  Rejected   Male    B  207
## 7  Admitted Female    B   17
## 8  Rejected Female    B    8
## 9  Admitted   Male    C  120
## 10 Rejected   Male    C  205
## 11 Admitted Female    C  202
## 12 Rejected Female    C  391
## 13 Admitted   Male    D  138
## 14 Rejected   Male    D  279
## 15 Admitted Female    D  131
## 16 Rejected Female    D  244
## 17 Admitted   Male    E   53
## 18 Rejected   Male    E  138
## 19 Admitted Female    E   94
## 20 Rejected Female    E  299
## 21 Admitted   Male    F   22
## 22 Rejected   Male    F  351
## 23 Admitted Female    F   24
## 24 Rejected Female    F  317
```

```r
calc_acceptance_rates <- function(df, department){
  df = df[df$Dept == department,]
  df = df[order(df$Admit),]
  df = df[order(df$Gender),]
  return(c(
    "Male" = df[1,4]/(df[1,4] + df[2,4]),
    "Female" = df[3,4]/(df[3,4] + df[4,4])
            ))
}

for(d in unique(ucb$Dept)){
  temp.admiss = (calc_acceptance_rates(ucb,d))
```

```r
  #print(temp.admiss)
  print(d)
  if(temp.admiss[[1]]>temp.admiss[[2]]){
    print("male higher")
  }else{
    print("female higher")
  }

}
```

```
## [1] "A"
## [1] "female higher"
## [1] "B"
## [1] "female higher"
## [1] "C"
## [1] "male higher"
## [1] "D"
## [1] "female higher"
## [1] "E"
## [1] "male higher"
## [1] "F"
## [1] "female higher"
```

The female acceptance rate is higher in 4 departments (A, B, D, F) and male acceptance rate is higher in 2 departments (C, E)

This can be considered a near example of simpson's paradox because our conclusion changes based on how we look at our data. In part A we looked at overall acceptance, but in part b we looked at the proportion of departments that accept male applicants at a higher rate.

Problem 4 (c)

Y – accepted? bool

X – m/f [1 = female]

Z – department – categorical

```r
calc_acceptance_rates <- function(df, department){
  df = df[df$Dept == department,]
  df = df[order(df$Admit),]
  df = df[order(df$Gender),]
  return(c(

    df[1,4]/(df[1,4] + df[2,4]),
    df[3,4]/(df[3,4] + df[4,4])

    ))
}

department.acceptances = data.frame("Gender" = c("Male","Female"))
for(d in unique(ucb$Dept)){

  #append(department.acceptances, c(calc_acceptance_rates(ucb,d)) )
```

```r
  department.acceptances[,paste("Z",d,sep = "=")] = calc_acceptance_rates(ucb,d)
}

department.acceptances
```

```
##   Gender       Z=A       Z=B       Z=C       Z=D       Z=E        Z=F
## 1   Male 0.6206061 0.6303571 0.3692308 0.3309353 0.2774869 0.05898123
## 2 Female 0.8240741 0.6800000 0.3406408 0.3493333 0.2391858 0.07038123
```

covariate - country of origin

treatment - drinks tea

outcome - if you live past 40

HW 3

1) a) $E[X] = \cancel{.5(0) + .5E[X|Z}$

$= .5E[X|Z=-1] + .5E[X|Z=1]$

$= .5(2) + .5(-2) = 0 = E[X]$

b) $V[X] = E[V[X|Z]] + V[E[X|Z]]$

$\cancel{= E[X^2] - E[E[X|Z]^2] + V[E[X|Z]]}$

$= .5(V[X|Z=-1]) + .5(V[X|Z=1]) + V[E[X|Z]]$

$= .5(4/3) + .5(4) + .5((2-0)^2 + (-2-0)^2)$

$= 8/3 + 4 = 20/3$

2) $\theta(x) = E[c(x)]$

~~$= E[c(x)]$~~

~~we know c(x) is the outcome for~~

$= E[c(x)|x]$ c(x) and x are independent

$= E[Y|x]$ given x we know c(x) is the outcome by definition of c(x)

$= r(x)$

3) a) $\theta = E[C_1] - E[C_0]$

$= (\tfrac{1}{2}(1) + \tfrac{1}{2}(0)) - (\tfrac{1}{2}(0) + \tfrac{1}{2}(1)) = \tfrac{1}{2} - \tfrac{1}{2}$

$\theta = 0$

b) $\alpha = E[Y \mid X=1] - E[Y \mid X=0]$

$= P(Y=1 \mid X=1) - P(Y=1 \mid X=0)$

$= P(Y=1 \mid X=1, Z=1) \cdot P(Z=1) + P(Y=1 \mid X=1, Z=0) \cdot P(Z=0) -$
$P(Y=1 \mid X=0, Z=1) \cdot P(Z=1) - P(Y=1 \mid X=0, Z=0) \cdot P(Z=0)$

$\cancel{= P(Y=1 \mid Z=1) \cdot P(Y=1 \mid Z=0) \cdot P(Y=1 \mid Z=1) - P(Y}$

$= P(Y=1 \mid Z=1) \cdot P(Z=1) + P(Y=1 \mid Z=0) \cdot P(Z=0) - P(Y=1 \mid Z=1) P(Z=1)$
$- P(Y=1 \mid Z=0) \cdot P(Z=0)$

$\cancel{= 1 \mid Z=0 \ 2 + -0 \ 1/2}$

$= 1 \cdot \tfrac{1}{2} + 0 \cdot \tfrac{1}{2} -$

b) $\alpha = E[Y|X=1] - E[Y|X=0]$

$$E[Y|X=1] = E[Y|X=1, Z=1] \cdot P(Z=1|X=1) + E[Y|X=1, Z=0] P(Z=0|X=1)$$

$$= E[C(1)|Z=1] \cdot P(Z=1|X=1) + E[C(1)|Z=0) P(Z=0|X=1)$$

$$= P[C(1)=1|Z=1) \cdot P(Z=1|X=1) + P(C(1)=1|Z=0) P(Z=0|X=1)$$

$$= P(Z=1|X=1)$$

$$E[Y|X=0] = E[Y|X=0, Z=1] \cdot P(Z=1|X=0) + E[Y|X=0, Z=0] \cdot P(Z=0|X=0)$$

$$= E[C(0)|Z=1] \cdot P(Z=1|X=0) + E[C(0)|Z=0] \cdot P(Z=0|X=0)$$

$$= P(C(0)=1|Z=1] \cdot P(Z=1|X=0) + P(C(0)=1|Z=0) \cdot P(Z=0|X=0)$$

$$= P(\cancel{Z=1|X=0}) + P(Z=0|X=0) = 1 - P(Z=1|X=0)$$

$$\alpha = P(Z=1|X=1) + P(Z=1|X=0) - 1$$

c) $\alpha = \underbrace{P(z=1|x=1)} + \underbrace{P(z=1|x=0)} - 1$

$$= \frac{P(x=1|z=1)\cdot P(z=1)}{P(x=1)}$$

$$= \frac{P(x=1|z=1)\cdot \cancel{P(z=1)}}{P(x=1|z=1)\cancel{P(z=1)} + P(x=1|z=0)\cdot \cancel{P(z=0)}} \qquad p(z=1) = p(z=0) = 1/2$$

$$= \frac{P(x=1|z=1)}{P(x=1|z=1) + P(x=1|z=0)}$$

$P(z=1|x=0)$

$$= \frac{P(x=0|z=1)\,P(z=1)}{P(x=0)}$$

$$= \frac{P(x=0|z=1)\,\cancel{P(z=1)}}{P(x=0|z=0)\cancel{P(z=0)} + P(x=0|z=1)\cancel{P(z=1)}}$$

$$= \frac{P(x=0|z=1)}{(1-P(x=1|z=0)) + (1-P(x=1|z=1))}$$

$$\alpha = \frac{P(x=1|z=1)}{P(x=1|z=1) + P(x=1|z=0)} + \frac{1-P(x=1|z=1)}{2-P(x=1|z=0)-P(x=1|z=1)} - 1$$

d) $\alpha = P(z=1|x=1) + P(z=1|x=0) - 1$

$= P(z=1) + P(z=1) - 1$ if we randomly assign x, then X and Z are independent

$= 0$

a practical example of this would be if our binary covariate ~~was~~ indicated whether someone lived in england or not, our treatment was drinking tea, and our outcome was whether the person lived past 40. If we don't randomly assign our treatment then we would find that people who drink tea live longer because english people tend to drink tea and also have a higher life expectancy. However, if we randomly assign people to drink tea or not then this ~~effect would~~ association would no longer exist because tea drinkers wouldn't be any more likely to be english.

e) if $P(x=1|z=1) = P(x=1|z=0) = .5$ the α would equal θ