

## exercise4

Xiang Li

2024/3/16

### Exercise 4.1

```
data_df = data.frame(X = c(0, 0, 0, 0, 1, 1, 1, 1), C = c(0, 0, 1, 1, 0, 0, 1, 1),  
  Y = c(0, 1, 0, 1, 0, 1, 0, 1), n = c(80, 20, 20, 10, 80, 20, 80, 40))  
data_df$Yxn = data_df$Y * data_df$n
```

a

```
EY1 = sum(data_df[data_df$X == 1, "Yxn"])/sum(data_df[data_df$X == 1, "n"])  
EY0 = sum(data_df[data_df$X == 0, "Yxn"])/sum(data_df[data_df$X == 0, "n"])  
print(c(EY1, EY0))
```

```
## [1] 0.2727273 0.2307692
```

```
print(EY1 - EY0)
```

```
## [1] 0.04195804
```

The confounder C makes the exchangeability assumption violated.

b

```
EY_CO_X0 = sum(data_df[(data_df$C == 0) & (data_df$X == 0), "Yxn"])/sum(data_df[(data_df$C ==  
  0) & (data_df$X == 0), "n"])  
print(EY_CO_X0)
```

```
## [1] 0.2
```

```
EY_C1_X0 = sum(data_df[(data_df$C == 1) & (data_df$X == 0), "Yxn"])/sum(data_df[(data_df$C ==  
  1) & (data_df$X == 0), "n"])  
print(EY_C1_X0)
```

```
## [1] 0.3333333
```

```
EY_CO_X1 = sum(data_df[(data_df$C == 0) & (data_df$X == 1), "Yxn"])/sum(data_df[(data_df$C ==
0) & (data_df$X == 1), "n"])
print(EY_CO_X1)
```

```
## [1] 0.2
```

```
EY_C1_X1 = sum(data_df[(data_df$C == 1) & (data_df$X == 1), "Yxn"])/sum(data_df[(data_df$C ==
1) & (data_df$X == 1), "n"])
print(EY_C1_X1)
```

```
## [1] 0.3333333
```

```
PC1 = sum(data_df[data_df$C == 1, "n"])/sum(data_df$n)
print(PC1)
```

```
## [1] 0.4285714
```

```
PC0 = sum(data_df[data_df$C == 0, "n"])/sum(data_df$n)
print(PC0)
```

```
## [1] 0.5714286
```

```
EY1 = EY_CO_X1 * PC0 + EY_C1_X1 * PC1
print(EY1)
```

```
## [1] 0.2571429
```

```
EY0 = EY_CO_X0 * PC0 + EY_C1_X0 * PC1
print(EY0)
```

```
## [1] 0.2571429
```

```
ATE = EY1 - EY0
print(ATE)
```

```
## [1] 0
```

```
PC0cX1 = sum(data_df[(data_df$C == 0) & (data_df$X == 1), "n"])/sum(data_df[data_df$X ==
1, "n"])
PC1cX1 = sum(data_df[(data_df$C == 1) & (data_df$X == 1), "n"])/sum(data_df[data_df$X ==
1, "n"])
EY1cX1 = EY_CO_X1 * PC0cX1 + EY_C1_X1 * PC1cX1
EY0cX1 = EY_CO_X0 * PC0cX1 + EY_C1_X0 * PC1cX1
ATT = EY1cX1 - EY0cX1
print(ATT)
```

```
## [1] 0
```

## Exercise 4.2

```
load("rhc_exercise.RData")
library(stdReg)
```

a

```
table(rhc$treatment, rhc$death30)
```

```
##
##      alive death
## no RHC  2463  1088
##  RHC    1354   830
```

b

```
outcome_reg = glm(death30 ~ treatment, data = rhc, family = "binomial")
print(summary(outcome_reg))
```

```
##
## Call:
## glm(formula = death30 ~ treatment, family = "binomial", data = rhc)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.81704    0.03640 -22.445  < 2e-16 ***
## treatmentRHC  0.32765    0.05717   5.731 9.98e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 7309.6  on 5734  degrees of freedom
## Residual deviance: 7276.8  on 5733  degrees of freedom
## AIC: 7280.8
##
## Number of Fisher Scoring iterations: 4
```

```
odds_ratio = exp(outcome_reg$coefficients[2])
print(odds_ratio)
```

```
## treatmentRHC
##      1.387698
```

No, it hasn't a causal interpretation, because there are confounders that cause assumption validated.

c

```
outcome_reg1 = glm(death30 ~ treatment + transhx + age + surv2md1 + scoma1 + hrt1 +
  bili1 + wtkilo1 + cat1 + aps1, data = rhc, family = "binomial")
print(outcome_reg1)
```

```
##
## Call:  glm(formula = death30 ~ treatment + transhx + age + surv2md1 +
##       scoma1 + hrt1 + bili1 + wtkilo1 + cat1 + aps1, family = "binomial",
##       data = rhc)
##
## Coefficients:
##           (Intercept)           treatmentRHC           transhx
##             -0.563190             0.217103             0.201820
##              age           surv2md1           scoma1
##             0.010097           -2.904026             0.003390
##             hrt1           bili1           wtkilo1
##             0.001974             0.057732           -0.003284
##            cat1CHF           cat1COPD           cat1Cirrhosis
##            -0.321003           -0.303914             0.347029
##      cat1Colon Cancer           cat1Coma      cat1Lung Cancer
##            -2.020822             0.972685             0.566511
## cat1MOSF w/Malignancy      cat1MOSF w/Sepsis           aps1
##             0.203666           -0.093597             0.009830
##
## Degrees of Freedom: 5734 Total (i.e. Null);  5717 Residual
## Null Deviance:      7310
## Residual Deviance: 6216  AIC: 6252
```

```
odds_ratio1 = exp(outcome_reg1$coefficients[2])
print(odds_ratio1)
```

```
## treatmentRHC
##      1.242472
```

The odds ratio increased after conditioning the confounding variables, which indicates that RHC increases the risk of death within 30 days.

**d**

```
outcome_reg2 = glm(death30 ~ treatment + transhx + age + surv2md1 + scoma1 + hrt1 +
  bili1 + wtkilo1 + cat1 + aps1 + treatment:transhx + treatment:age + treatment:surv2md1 +
  treatment:scoma1 + treatment:hrt1 + treatment:bili1 + treatment:wtkilo1 + treatment:cat1 +
  treatment:aps1, data = rhc, family = "binomial")
print(outcome_reg2)
```

```
##
## Call:  glm(formula = death30 ~ treatment + transhx + age + surv2md1 +
##       scoma1 + hrt1 + bili1 + wtkilo1 + cat1 + aps1 + treatment:transhx +
##       treatment:age + treatment:surv2md1 + treatment:scoma1 + treatment:hrt1 +
##       treatment:bili1 + treatment:wtkilo1 + treatment:cat1 + treatment:aps1,
```

```
## family = "binomial", data = rhc)
##
## Coefficients:
## (Intercept) treatmentRHC
## -0.0055900 -1.5695237
## transhx age
## 0.4099810 0.0076201
## surv2md1 scoma1
## -3.5860788 0.0045004
## hrt1 bili1
## 0.0020014 0.0626420
## wtkilo1 cat1CHF
## -0.0016927 -0.3122862
## cat1COPD cat1Cirrhosis
## -0.2379022 0.2719073
## cat1Colon Cancer cat1Coma
## -2.2383314 0.7993026
## cat1Lung Cancer cat1MOSF w/Malignancy
## 0.7091873 0.0819996
## cat1MOSF w/Sepsis aps1
## -0.2228356 0.0070868
## treatmentRHC:transhx treatmentRHC:age
## -0.4681405 0.0085788
## treatmentRHC:surv2md1 treatmentRHC:scoma1
## 1.9502776 -0.0022476
## treatmentRHC:hrt1 treatmentRHC:bili1
## -0.0004042 -0.0100374
## treatmentRHC:wtkilo1 treatmentRHC:cat1CHF
## -0.0045461 -0.0508556
## treatmentRHC:cat1COPD treatmentRHC:cat1Cirrhosis
## -0.1452940 0.2314418
## treatmentRHC:cat1Colon Cancer treatmentRHC:cat1Coma
## -7.9434049 0.3066561
## treatmentRHC:cat1Lung Cancer treatmentRHC:cat1MOSF w/Malignancy
## -1.5474373 0.3587848
## treatmentRHC:cat1MOSF w/Sepsis treatmentRHC:aps1
## 0.2331030 0.0099060
##
## Degrees of Freedom: 5734 Total (i.e. Null); 5701 Residual
## Null Deviance: 7310
## Residual Deviance: 6181 AIC: 6249
```

e

```
rhcl = rhc
rhcl$treatment = "RHC"
prob_y1 = predict(outcome_reg2, newdata = rhcl, type = "response")
```

**f**

```
EY1 = mean(prob_y1)
print(EY1)
```

```
## [1] 0.3570939
```

$E(Y(1)) = 0.357$ . The average risk of death is 0.357 if every patient receives RHC treatment.

**g**

```
rhc0 = rhc
rhc0$treatment = "no RHC"
prob_y0 = predict(outcome_reg2, newdata = rhc0, type = "response")
EY0 = mean(prob_y0)
print(EY0)
```

```
## [1] 0.3187898
```

$E(Y(0)) = 0.319$ . The average risk of death is 0.319 if every patient doesn't receive RHC treatment.

**h**

```
ATE = EY1 - EY0
print(ATE)
```

```
## [1] 0.03830408
```

```
odd_ratio2 = (EY1/(1 - EY1))/(EY0/(1 - EY0))
print(odd_ratio2)
```

```
## [1] 1.186893
```

The ATE is 0.038, means that receiving RHC treatment will increase 3.8% risk of death compared with not receiving RHC treatment.

The odds ratio is 1.187.

**i**

```
print(odds_ratio1 - odd_ratio2)
```

```
## treatmentRHC
## 0.05557914
```

j

```
std = stdGlm(fit = outcome_reg2, data = rhc, X = "treatment")
print(summary(std))
```

```
##
## Formula: death30 ~ treatment + transhx + age + surv2md1 + scoma1 + hrt1 +
##      bili1 + wtkilo1 + cat1 + aps1 + treatment:transhx + treatment:age +
##      treatment:surv2md1 + treatment:scoma1 + treatment:hrt1 +
##      treatment:bili1 + treatment:wtkilo1 + treatment:cat1 + treatment:aps1
## Family: binomial
## Link function: logit
## Exposure:  treatment
##
##      Estimate Std. Error lower 0.95 upper 0.95
## no RHC      0.319      0.00993      0.299      0.338
## RHC         0.357      0.00868      0.340      0.374
```

$E(Y(1))$  is 0.357 and  $E(Y(0))$  is 0.319.

```
print(summary(std, contrast = "difference", reference = "no RHC"))
```

```
##
## Formula: death30 ~ treatment + transhx + age + surv2md1 + scoma1 + hrt1 +
##      bili1 + wtkilo1 + cat1 + aps1 + treatment:transhx + treatment:age +
##      treatment:surv2md1 + treatment:scoma1 + treatment:hrt1 +
##      treatment:bili1 + treatment:wtkilo1 + treatment:cat1 + treatment:aps1
## Family: binomial
## Link function: logit
## Exposure:  treatment
## Reference level: treatment = no RHC
## Contrast:  difference
##
##      Estimate Std. Error lower 0.95 upper 0.95
## no RHC      0.0000      0.0000      0.0000      0.0000
## RHC         0.0383      0.0127      0.0135      0.0632
```

ATE is 0.038 and 95% confidence interval of ATE is (0.013, 0.064).

k

```
std1 = stdGlm(fit = outcome_reg2, data = rhc, X = "treatment", subsetnew = treatment ==
              "RHC")
print(summary(std1))
```

```
##
## Formula: death30 ~ treatment + transhx + age + surv2md1 + scoma1 + hrt1 +
##      bili1 + wtkilo1 + cat1 + aps1 + treatment:transhx + treatment:age +
##      treatment:surv2md1 + treatment:scoma1 + treatment:hrt1 +
```

```
##      treatment:bili1 + treatment:wtkilo1 + treatment:cat1 + treatment:aps1
## Family: binomial
## Link function: logit
## Exposure:  treatment
##
##      Estimate Std. Error lower 0.95 upper 0.95
## no RHC      0.339      0.00977      0.320      0.358
## RHC          0.380      0.01129      0.358      0.402
```

$E(Y(1) \mid \text{treatment} = \text{RHC})$  is 0.339 and  $E(Y(0) \mid \text{treatment} = \text{RHC})$  is 0.380.

```
print(summary(std1, contrast = "difference", reference = "no RHC"))
```

```
##
## Formula: death30 ~ treatment + transhx + age + surv2md1 + scoma1 + hrt1 +
##      bili1 + wtkilo1 + cat1 + aps1 + treatment:transhx + treatment:age +
##      treatment:surv2md1 + treatment:scoma1 + treatment:hrt1 +
##      treatment:bili1 + treatment:wtkilo1 + treatment:cat1 + treatment:aps1
## Family: binomial
## Link function: logit
## Exposure:  treatment
## Reference level:  treatment = no RHC
## Contrast:  difference
##
##      Estimate Std. Error lower 0.95 upper 0.95
## no RHC      0.0000      0.0000      0.0000      0.0000
## RHC          0.0411      0.0137      0.0143      0.0679
```

ATT is 0.0411 and 95% confidence interval of ATE is (0.0145, 0.0676).

For those who currently receive RHC treatment, 4.11% less would have died if they would have not received RHC treatment.