Homework 8

Xinyi Lin

11/17/2019

library(tidyverse)  
library(survey)

hw8\_data = read.csv("./data\_hw8.csv")  
head(hw8\_data)

## X id t0 L1 L2 L3 A Y  
## 1 1 1 0 3 0.5774203 8 1 NA  
## 2 2 1 1 3 -1.6196027 8 1 NA  
## 3 3 1 2 3 -1.3912145 8 1 NA  
## 4 4 1 3 2 -1.3814392 8 0 NA  
## 5 5 1 4 1 0.4641491 8 1 NA  
## 6 6 1 5 3 -1.5809908 8 1 NA

## Question 1

1. Time point 1: L1, L2, L3, A
2. Time point 2: L1, L2, A
3. Time point 3: L1, L2, A
4. Time point 4: L1, L2, A
5. Time point 5: L1, L2, A
6. Time point 6: L1, L2, A, Y

## Question 2

The causal contrast is:

, for .

## Question 3

Let indicates covariate in time point .

## Question 4

1. No unmeasured exposure-outcome confounding given C
2. No unmeasured mediator-outcome confounding given C
3. No unmeasured exposure-mediator confounding given C
4. No effect of exposure that confounds the mediator-outcome relationship

## Question 5

# create wide data  
AC\_data = hw8\_data %>%   
 gather(key = "Lt", value = "value", c(L1,L2,A)) %>%   
 arrange(id) %>%   
 mutate(Lt = str\_c(Lt, t0)) %>%   
 select(id, Lt, value) %>%   
 spread(key = Lt, value = value)  
Y\_data = hw8\_data %>%   
 select(id, L3, Y) %>%   
 na.omit()  
wide\_data = merge(AC\_data, Y\_data)

# Time point 0  
glm.model0 = glm(A0~L3, data = wide\_data, family = binomial)  
p0 = predict(glm.model0, type = "response")  
w0 = ifelse(wide\_data$A0==1, 1/p0, 1/(1-p0))  
# Time point 1  
glm.model1 = glm(A1~L3+A0+L10+L20, data = wide\_data, family = binomial)  
p1 = predict(glm.model1, type = "response")  
w1 = ifelse(wide\_data$A1==1, 1/p1, 1/(1-p1))  
# Time point 2  
glm.model2 = glm(A2~A1+L11+L21+L3+A0+L10+L20, data = wide\_data, family = binomial)  
p2 = predict(glm.model2, type = "response")  
w2 = ifelse(wide\_data$A2==1, 1/p2, 1/(1-p2))  
# Time point 3  
glm.model3 = glm(A3~A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = wide\_data, family = binomial)  
p3 = predict(glm.model3, type = "response")  
w3 = ifelse(wide\_data$A3==1, 1/p3, 1/(1-p3))  
# Time point 4  
glm.model4 = glm(A4~A3+L13+L23+A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = wide\_data, family = binomial)  
p4 = predict(glm.model4, type = "response")  
w4 = ifelse(wide\_data$A4==1, 1/p4, 1/(1-p4))  
# Time point 5  
glm.model5 = glm(A5~A4+L14+L24+A3+L13+L23+A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = wide\_data, family = binomial)  
p5 = predict(glm.model5, type = "response")  
w5 = ifelse(wide\_data$A5==1, 1/p5, 1/(1-p5))  
# Time point 6  
glm.model6 = glm(A6~A5+L15+L25+A4+L14+L24+A3+L13+L23+A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = wide\_data, family = binomial)  
p6 = predict(glm.model6, type = "response")  
w6 = ifelse(wide\_data$A6==1, 1/p6, 1/(1-p6))  
w = w0\*w1\*w2\*w3\*w4\*w5\*w6

# Question 6

Marginal model:

set.seed(123)  
nboots = 1000  
n\_sample = nrow(wide\_data)  
  
beta = rep(NA, nboots)  
beta0 = rep(NA, nboots)  
beta1 = rep(NA, nboots)  
beta2 = rep(NA, nboots)  
beta3 = rep(NA, nboots)  
beta4 = rep(NA, nboots)  
beta5 = rep(NA, nboots)  
beta6 = rep(NA, nboots)  
  
for (i in 1:nboots) {  
 S.b <- sample(1:n\_sample, size = n\_sample, replace = TRUE)  
 data.b <- wide\_data[S.b, ]  
 # Time point 0  
 glm.model0 = glm(A0~L3, data = data.b, family = binomial)  
 p0 = predict(glm.model0, type = "response")  
 w0 = ifelse(data.b$A0==1, 1/p0, 1/(1-p0))  
 # Time point 1  
 glm.model1 = glm(A1~L3+A0+L10+L20, data = data.b, family = binomial)  
 p1 = predict(glm.model1, type = "response")  
 w1 = ifelse(data.b$A1==1, 1/p1, 1/(1-p1))  
 # Time point 2  
 glm.model2 = glm(A2~A1+L11+L21+L3+A0+L10+L20, data = data.b, family = binomial)  
 p2 = predict(glm.model2, type = "response")  
 w2 = ifelse(data.b$A2==1, 1/p2, 1/(1-p2))  
 # Time point 3  
 glm.model3 = glm(A3~A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = data.b, family = binomial)  
 p3 = predict(glm.model3, type = "response")  
 w3 = ifelse(data.b$A3==1, 1/p3, 1/(1-p3))  
 # Time point 4  
 glm.model4 = glm(A4~A3+L13+L23+A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = data.b, family = binomial)  
 p4 = predict(glm.model4, type = "response")  
 w4 = ifelse(data.b$A4==1, 1/p4, 1/(1-p4))  
 # Time point 5  
 glm.model5 = glm(A5~A4+L14+L24+A3+L13+L23+A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = data.b, family = binomial)  
 p5 = predict(glm.model5, type = "response")  
 w5 = ifelse(data.b$A5==1, 1/p5, 1/(1-p5))  
 # Time point 6  
 glm.model6 = glm(A6~A5+L15+L25+A4+L14+L24+A3+L13+L23+A2+L12+L22+A1+L11+L21+L3+A0+L10+L20, data = data.b, family = binomial)  
 p6 = predict(glm.model6, type = "response")  
 w6 = ifelse(data.b$A6==1, 1/p6, 1/(1-p6))  
 w = w0\*w1\*w2\*w3\*w4\*w5\*w6  
   
 data.b$w = w  
 design = svydesign(ids = ~id, weights = ~w, data = data.b)  
 msm = svyglm(Y ~ A0 + A1 + A2 + A3 + A4 + A5 + A6, family = gaussian(link = "identity"), design = design)  
 beta[i] = msm$coef[1]  
 beta0[i] = msm$coef[2]  
 beta1[i] = msm$coef[3]  
 beta2[i] = msm$coef[4]  
 beta3[i] = msm$coef[5]  
 beta4[i] = msm$coef[6]  
 beta5[i] = msm$coef[7]  
 beta6[i] = msm$coef[8]  
}

beta\_est = mean(beta)  
CIL\_beta = mean(beta) - 1.96\*sqrt(var(beta))  
CIU\_beta = mean(beta) + 1.96\*sqrt(var(beta))  
beta0\_est = mean(beta0)  
CIL\_beta0 = mean(beta0) - 1.96\*sqrt(var(beta0))  
CIU\_beta0 = mean(beta0) + 1.96\*sqrt(var(beta0))  
beta1\_est = mean(beta1)  
CIL\_beta1 = mean(beta1) - 1.96\*sqrt(var(beta1))  
CIU\_beta1 = mean(beta1) + 1.96\*sqrt(var(beta1))  
beta2\_est = mean(beta2)  
CIL\_beta2 = mean(beta2) - 1.96\*sqrt(var(beta2))  
CIU\_beta2 = mean(beta2) + 1.96\*sqrt(var(beta2))  
beta3\_est = mean(beta3)  
CIL\_beta3 = mean(beta3) - 1.96\*sqrt(var(beta3))  
CIU\_beta3 = mean(beta3) + 1.96\*sqrt(var(beta3))  
beta4\_est = mean(beta4)  
CIL\_beta4 = mean(beta4) - 1.96\*sqrt(var(beta4))  
CIU\_beta4 = mean(beta4) + 1.96\*sqrt(var(beta4))  
beta5\_est = mean(beta5)  
CIL\_beta5 = mean(beta5) - 1.96\*sqrt(var(beta5))  
CIU\_beta5 = mean(beta5) + 1.96\*sqrt(var(beta5))  
beta6\_est = mean(beta6)  
CIL\_beta6 = mean(beta6) - 1.96\*sqrt(var(beta6))  
CIU\_beta6 = mean(beta6) + 1.96\*sqrt(var(beta6))

The estimated value of is -2.25 and 95% confidence interval is (-2.706, -1.794).

The estimated value of is -0.052 and 95% confidence interval is (-0.231, 0.128).

The estimated value of is 0.107 and 95% confidence interval is (-0.05, 0.264).

The estimated value of is -0.213 and 95% confidence interval is (-0.401, -0.026).

The estimated value of is 0.049 and 95% confidence interval is (-0.107, 0.205).

The estimated value of is -0.048 and 95% confidence interval is (-0.284, 0.188).

The estimated value of is -0.329 and 95% confidence interval is (-0.5, -0.157).

The estimated value of is -2.231 and 95% confidence interval is (-2.398, -2.064).

## Question 7

Interpretation:

: The estimated value of outcome on average if the subject is always in control group is -2.25 and we have 95% confidence that the ture value lies between -2.706 and -1.794.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 0 is -0.052 and we have 95% confidence that the ture value lies between -0.231 and 0.128.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 1 is 0.107 and we have 95% confidence that the ture value lies between -0.05 and 0.264.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 2 is -0.213 and we have 95% confidence that the ture value lies between -0.401 and -0.026.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 3 is 0.049 and we have 95% confidence that the ture value lies between -0.107 and 0.205.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 4 is -0.048 and we have 95% confidence that the ture value lies between -0.284 and 0.188.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 5 is -0.329 and we have 95% confidence that the ture value lies between -0.5 and -0.157.

: On average, the difference of estimated value of outcome between treatment group and control group in time point 6 is -2.231 and we have 95% confidence that the ture value lies between -2.398 and -2.064.

## Question 8

Assumptions:

1. Consistency;
2. Stable Unit Treatment Value Assumption(SUTVA);
3. Exchangeability;
4. Positivity.

## Question 9

According to estimated values of coefficients and 95% CI, we can find that are significant, which means treatments in time point 2, 5, 6 bring significant difference in outcome.