hw8

uni:xw2598

2019/11/16

1) (10 points) Look at the data. List all the variables of interest measured at each time point.

```
df <- read_csv("data_hw8.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   id = col_integer(),
##   t0 = col_integer(),
##   L1 = col_integer(),
##   L2 = col_double(),
##   L3 = col_integer(),
##   A = col_integer(),
##   Y = col_double()
## )
```

```
df$L1 = as.factor(df$L1)
df$A = as.factor(df$A)
vars = c("L1" , "L2", "L3","A" )

tab <- CreateTableOne(vars = vars, strata = "t0", data = df, test = FALSE)
print(tab, smd = TRUE)
```

```
##                 Stratified by t0
##                  0            1            2            3
##   n               2500          2500          2500          2500
##   L1 (%)
##      1             805 (32.2)    515 (20.6)    370 (14.8)    387 (15.5)
##      2             870 (34.8)    717 (28.7)   1198 (47.9)   1184 (47.4)
##      3             825 (33.0)   1268 (50.7)    932 (37.3)    929 (37.2)
##   L2 (mean (SD)) 0.50 (0.10)  -1.11 (0.79)  -1.17 (0.75)  -1.15 (0.77)
##   L3 (mean (SD)) 7.45 (1.69)   7.45 (1.69)   7.45 (1.69)   7.45 (1.69)
##   A = 1 (%)      2022 (80.9)   2085 (83.4)   2065 (82.6)   2064 (82.6)
##                 Stratified by t0
##                  4             5             6             SMD
##   n               2500          2500          2500
##   L1 (%)                                                    0.237
##      1             353 (14.1)    355 (14.2)    350 (14.0)
##      2            1210 (48.4)   1212 (48.5)   1256 (50.2)
##      3             937 (37.5)    933 (37.3)    894 (35.8)
##   L2 (mean (SD)) -1.15 (0.76)  -1.17 (0.76)  -1.16 (0.76)   0.884
##   L3 (mean (SD))  7.45 (1.69)   7.45 (1.69)   7.45 (1.69)  <0.001
##   A = 1 (%)       2078 (83.1)   2073 (82.9)   2051 (82.0)   0.025
```

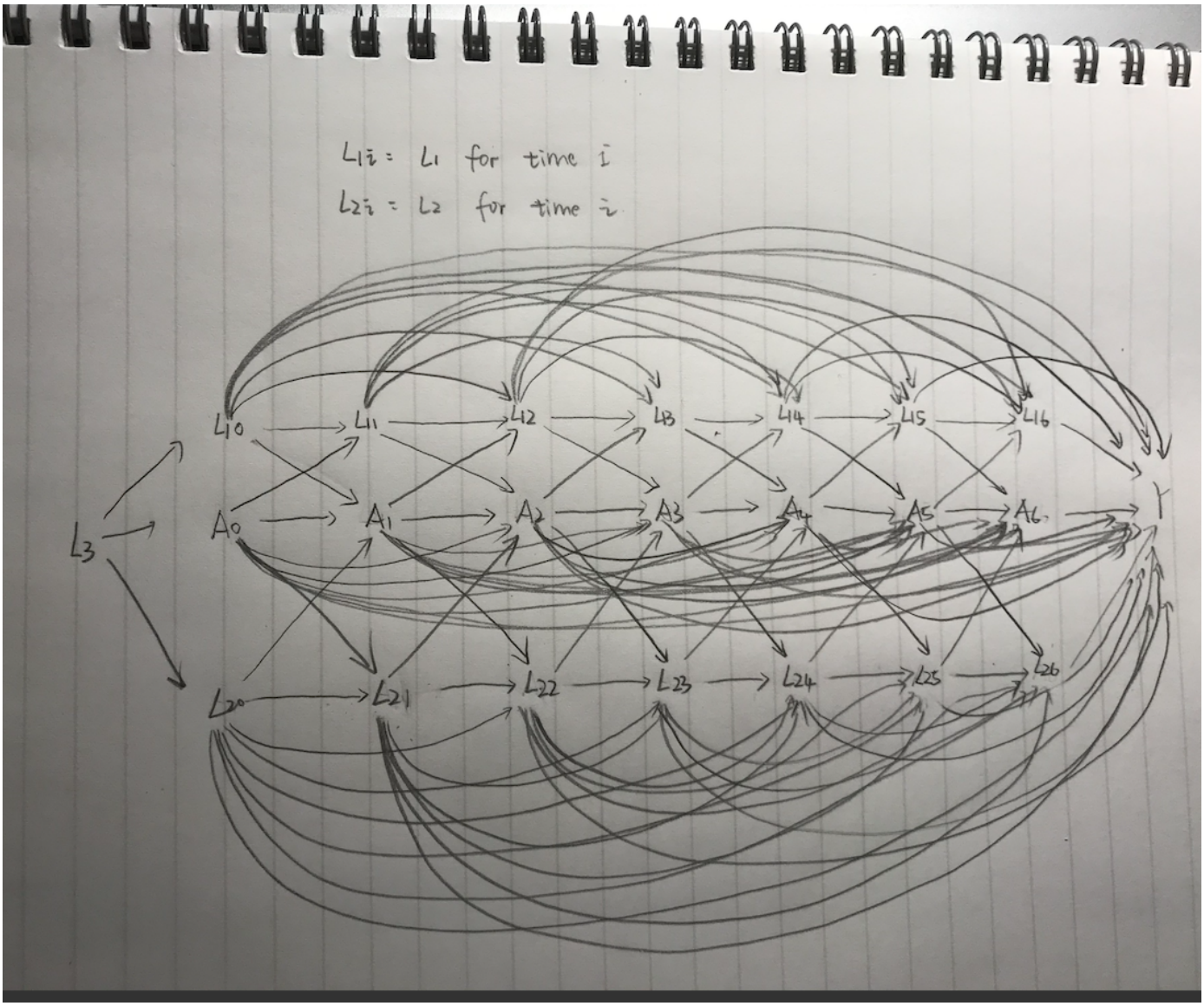
The interested variables: L1,L2,L3,A. L3 is a baseline corvariate. L2,L3, A will be recorded at each time point.

2) (10 points) Write the causal contrast of interest.

$$ACE = E[Y_{A_0=1,A_1=1,A_2=1,A_3=1,A_4=1,A_5=1,A_6=1} - Y_{A_0=0,A_1=0,A_2=0,A_3=0,A_4=0,A_5=0,A_6=0}]$$

We are interested in the difference between Y from “never treat” group and Y from “always treat” group during the 7 time points.

3) (10 points) Write the complete causal DAG for this longitudinal study assuming that it arises from either an observational study or a blocked and adaptive randomized experiment.



4) (15 points) List the set of no unmeasured confounding assumptions that we need to satisfy in order to identify the causal contrast of interest.

$$Y_{a0a1a2a3a4a5a6} \perp A_0 | L_3$$

$$Y_{a0a1a2a3a4a5a6} \perp A_1 | \{L_{10}, A_0, L_{20}, L_3\}$$

$$Y_{a0a1a2a3a4a5a6} \perp A_2 | \{L_{10}, A_0, L_{20}, L_{11}, A_1, L_{21}, L_3\}$$

$$Y_{a0a1a2a3a4a5a6} \perp A_3 | \{L_{10}, A_0, L_{20}, L_{11}, A_1, L_{21}, L_{12}, A_2, L_{22}, L_3\}$$

$$Y_{a0a1a2a3a4a5a6} \perp A_4 | \{L_{10}, A_0, L_{20}, L_{11}, A_1, L_{21}, L_{12}, A_2, L_{22}, L_{13}, A_3, L_{23}, L_3\}$$

$$Y_{a0a1a2a3a4a5a6} \perp A_5 | \{L_{10}, A_0, L_{20}, L_{11}, A_1, L_{21}, L_{12}, A_2, L_{22}, L_{13}, A_3, L_{23}, L_{14}, A_4, L_{24}, L_3\}$$

$$Y_{a0a1a2a3a4a5a6} \perp A_6 | \{L_{10}, A_0, L_{20}, L_{11}, A_1, L_{21}, L_{12}, A_2, L_{22}, L_{13}, A_3, L_{23}, L_{14}, A_4, L_{24}, L_{15}, A_5, L_{25}, L_3\}$$

5) (15 points) Fit the propensity score models required to recover the pseudo population where treated and untreated groups at each time point are exchangeable.

```
df2 = df %>% select(-X1,-Y)
t0 = filter(df2,t0==0)
t1 = filter(df2,t0==1)
df3 <- merge(t0,t1,by=c("id"),suffix = c('0','1'))
t2 = filter(df2,t0==2)
df3 <- merge(df3,t2,by=c("id"),suffix = c('1','2'))
t3 = filter(df2,t0==3)
df3 <- merge(df3,t3,by=c("id"),suffixes = c("2","3"))
t4 = filter(df2,t0==4)
df3 <- merge(df3,t4,by=c("id"),suffixes = c("3","4"))
t5 = filter(df2,t0==5)
df3 <- merge(df3,t5,by=c("id"),suffixes = c("4","5"))
t6 = filter(df2,t0==6)
df3 <- merge(df3,t6,by=c("id"),suffixes = c("5","6"))
v = names(df3)
v[32:36] = c("t06", "L16", "L26", "L36", "A6" )
colnames(df3) = v
df3 = df3 %>% dplyr::select(-L31,-L32,-L33,-L34,-L35,-L36)

ps.a0 <- glm(A0~L30 ,data=df3, family = binomial)
ps.a1 <- glm(A1~L30+L20+L10+A0 ,data=df3, family = binomial)
ps.a2 <- glm(A2~L30+L20+L10+A0+A1+L21+L11 ,data=df3, family = binomial)
ps.a3 <- glm(A3~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12 ,data=df3, family = binomial)
ps.a4 <- glm(A4~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12+A3+L13+L23 ,data=df3, family = b
inomial)
ps.a5 <- glm(A5~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12+A3+L13+L23+A4+L14+L24 ,data=df3,
family = binomial)
ps.a6 <- glm(A6~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12+A3+L13+L23+A4+L14+L24+A5+L15+L25
,data=df3, family = binomial)
```

6) (10 points) Write the model specification and fit the marginal structural model. Show the estimates and 95% confidence intervals for all parameters.

The model:

$$E[Y_{a0a1a2a3a4a5a6}] = \beta_0 + \beta_1 A0 + \beta_2 A1 + \beta_3 A2 + \beta_4 A3 + \beta_5 A4 + \beta_6 A5 + \beta_7 A6$$

```
###
nboot <- 1000
n <- nrow(df3)

a0.holder <- rep(NA)
a1.holder <-rep(NA)
a2.holder <- rep(NA)
a3.holder <- rep(NA)
a4.holder <- rep(NA)
a5.holder <- rep(NA)
a6.holder <- rep(NA)
est.w <-rep(NA)
Y = filter(df,is.na(Y)==FALSE) %>% select(Y)
df4 = data.frame(df3,Y)

for (b in 1:nboot){
  set.seed(123+b)
  S.b <- sample(1:n, size = n, replace = TRUE)
  data.b <- df4[S.b, ]

ps.a0 <- glm(A0~L30 ,data=data.b, family = binomial)
ps.a1 <- glm(A1~L30+L20+L10+A0 ,data=data.b, family = binomial)
ps.a2 <- glm(A2~L30+L20+L10+A0+A1+L21+L11 ,data=data.b, family = binomial)
ps.a3 <- glm(A3~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12 ,data=data.b, family = binomial)
ps.a4 <- glm(A4~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12+A3+L13+L23 ,data=data.b, family
 = binomial)
ps.a5 <- glm(A5~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12+A3+L13+L23+A4+L14+L24 ,data=dat
a.b, family = binomial)
ps.a6 <- glm(A6~L30+L20+L10+A0+A1+L21+L11+A2+L22+L12+A3+L13+L23+A4+L14+L24+A5+L15+L25
,data=data.b, family = binomial)

psa0 <- predict(ps.a0, type="response")
psa1 <- predict(ps.a1, type="response")
psa2 <- predict(ps.a2, type="response")
psa3 <- predict(ps.a3, type="response")
psa4 <- predict(ps.a4, type="response")
psa5 <- predict(ps.a5, type="response")
psa6 <- predict(ps.a6, type="response")


  est.w = 1*ifelse(df3$A0==1,1/psa0,1/(1-psa0))
  est.w = est.w*ifelse(df3$A1==1,1/psa1,1/(1-psa1))
 est.w = est.w*ifelse(df3$A2==1,1/psa2,1/(1-psa2))
 est.w = est.w*ifelse(df3$A3==1,1/psa3,1/(1-psa3))
 est.w = est.w*ifelse(df3$A4==1,1/psa4,1/(1-psa4))
 est.w = est.w*ifelse(df3$A5==1,1/psa5,1/(1-psa5))
 est.w = est.w*ifelse(df3$A6==1,1/psa6,1/(1-psa6))


msm <-svyglm(Y~A0+A1+A2+A3+A4+A5+A6,design = svydesign(~ 1, weights = ~ est.w,data=da
ta.b))

a0.holder[b] <- msm$coef[2]
a1.holder[b] <- msm$coef[3]
a2.holder[b] <- msm$coef[4]
a3.holder[b] <- msm$coef[5]
a4.holder[b] <- msm$coef[6]
```

```
a5.holder[b] <- msm$coef[7]
a6.holder[b] <- msm$coef[8]

}


data.frame(a0 = mean(a0.holder),a1 = mean(a1.holder),a2 = mean(a2.holder),a3 = mean(a
3.holder),a4 = mean(a4.holder),a5 = mean(a5.holder),a6 = mean(a6.holder),row.names =
"estimate")
```

```
##                   a0          a1           a2         a3         a4
## estimate -0.0166245 -0.02184553 -0.001232177 0.01977068 0.06940319
##                   a5        a6
## estimate -0.2188297 -2.197512
```

```
cil = function(x){
  mean(x)-1.96*sd(x)
}
ciu = function(x){
  mean(x)+1.96*sd(x)
}
data.frame(a0 = mean(a0.holder),a1 = mean(a1.holder),a2 = mean(a2.holder),a3 = mean(a
3.holder),a4 = mean(a4.holder),a5 = mean(a5.holder),a6 = mean(a6.holder),row.names =
"estimate")
```

```
##                   a0          a1           a2         a3         a4
## estimate -0.0166245 -0.02184553 -0.001232177 0.01977068 0.06940319
##                   a5        a6
## estimate -0.2188297 -2.197512
```

```
 data.frame(a0 = quantile(a0.holder,probs = 0.025),a1 = quantile(a1.holder,probs = 0.
025),a2 = quantile(a2.holder,probs = 0.025),a3 = quantile(a3.holder,probs = 0.025),a4
= quantile(a4.holder,probs = 0.025),a5 = quantile(a5.holder,probs = 0.025),a6 = quant
ile(a6.holder,probs = 0.025),row.names = "lower bound of 95%CI")
```

```
##                               a0         a1         a2         a3
## lower bound of 95%CI -0.3431468 -0.3190305 -0.2851383 -0.2597972
##                               a4         a5        a6
## lower bound of 95%CI -0.2568132 -0.5421722 -2.514993
```

```
 data.frame(a0 = quantile(a0.holder,probs = 0.975),a1 = quantile(a1.holder,probs = 0.
975),a2 = quantile(a2.holder,probs = 0.975),a3 = quantile(a3.holder,probs = 0.975),a4
= quantile(a4.holder,probs = 0.975),a5 = quantile(a5.holder,probs = 0.975),a6 = quant
ile(a6.holder,probs = 0.975),row.names = "upper bound of 95%CI")
```

```
##                              a0        a1        a2        a3        a4
## upper bound of 95%CI 0.278725 0.2781602 0.2700232 0.3201236 0.3534568
##                               a5        a6
## upper bound of 95%CI 0.09302849 -1.880957
```

7) (10 points) Give an associational interpretation to all the parameters estimated with the marginal structural model.

A hypothetical intervention to change from control group at baseline to treatment group would decrease the outcome by about 0.0166.

A hypothetical intervention to change from control group to treatment group at time point 1 would decrease the outcome by about 0.02184553.

A hypothetical intervention to change from control group to treatment group at time point 2 would decrease the outcome by about 0.001232177.

A hypothetical intervention to change from control group to treatment group at time point 3 would increase the outcome by about 0.01977068.

A hypothetical intervention to change from control group to treatment group at time point 4 would increase the outcome by about 0.06940319 .

A hypothetical intervention to change from control group to treatment group at time point 5 would decrease the outcome by about 0.2188297.

A hypothetical intervention to change from control group to treatment group at time point 6 would decrease the outcome by about 2.197512

8) (10 points) List the assumptions under which the parameters estimated with the marginal structural model have a causal interpretation.

1. no unmeasured confounding assumptions (2)positivity (3)exchangeability (4)stable unit treatment value assumption (5)consistency

9) (10 points) Give a causal interpretation to all the parameters estimated with the marginal structural model. What do you conclude?

An intervention to change outcome at baseline would have a negative causal effect which indicates the difference of the outcome of treatment group and control group at baseline is -0.0166.

An intervention to change outcome at time point 1 would have a negative causal effect which indicates the difference of the outcome of treatment group and control group at time point 1 is -0.02184553.

An intervention to change outcome at time point 2 would have a negative causal effect which indicates the difference of the outcome of treatment group and control group at time point 2 is 0.001232177.

An intervention to change outcome at time point 3 would have a positive causal effect which indicates the difference of the outcome of treatment group and control group at time point 3 is 0.01977068.

An intervention to change outcome at time point 4 would have a positive causal effect which indicates the difference of the outcome of treatment group and control group at time point 4 is 0.06940319

An intervention to change outcome at time point 5 would have a negative causal effect which indicates the difference of the outcome of treatment group and control group at time point 5 is -0.2188297.

An intervention to change outcome at time point 6 would have a negative causal effect which indicates the difference of the outcome of treatment group and control group at time point 6 is -2.197512

Conclusion:

Only when we make the assumptions above to meet the requirements of causal inference analysiscan we draw causal conclusion, which is different from associational interpretation. associational interpretation is based on the observed data while causal inference is mainly on assumptions and inference on unobserved data.