Homework 6

Yuki Joyama

```
# libraries
library(tidyverse)
library(ggplot2)
library(readr)
library(dagitty)
library(knitr)

# install.packages("remotes")
# remotes::install_github("LindaValeri/CMAverse")
library(CMAverse)

# setup plot theme
theme_set(
  theme_minimal() +
    theme(legend.position = "top")
  )
```

```
# import data
df = read_csv("./data/sim_data_bangladesh.csv") |>
  dplyr::select(-"...1") |>
  mutate(
    female = as.factor(female)
  )
```

Hypothesis: fetal growth could be involved in the mechanism explaining the effect of manganese on child neurodevelopment

1. NDE, NIE, TE

Exposure: `ln_mn_c`
Mediator: `birthlength_c`
Outcome: `cognitive_raw`
Covariates: `female`, `approxage`, `protein_c`

I will estimate causal effects using the following code:

```
med_result = cmest(
  data = df,
  model = "rb", # regression-based approach
  outcome = "cognitive_raw",
  exposure = "ln_mn_c",
  mediator = "birthlength_c",
```

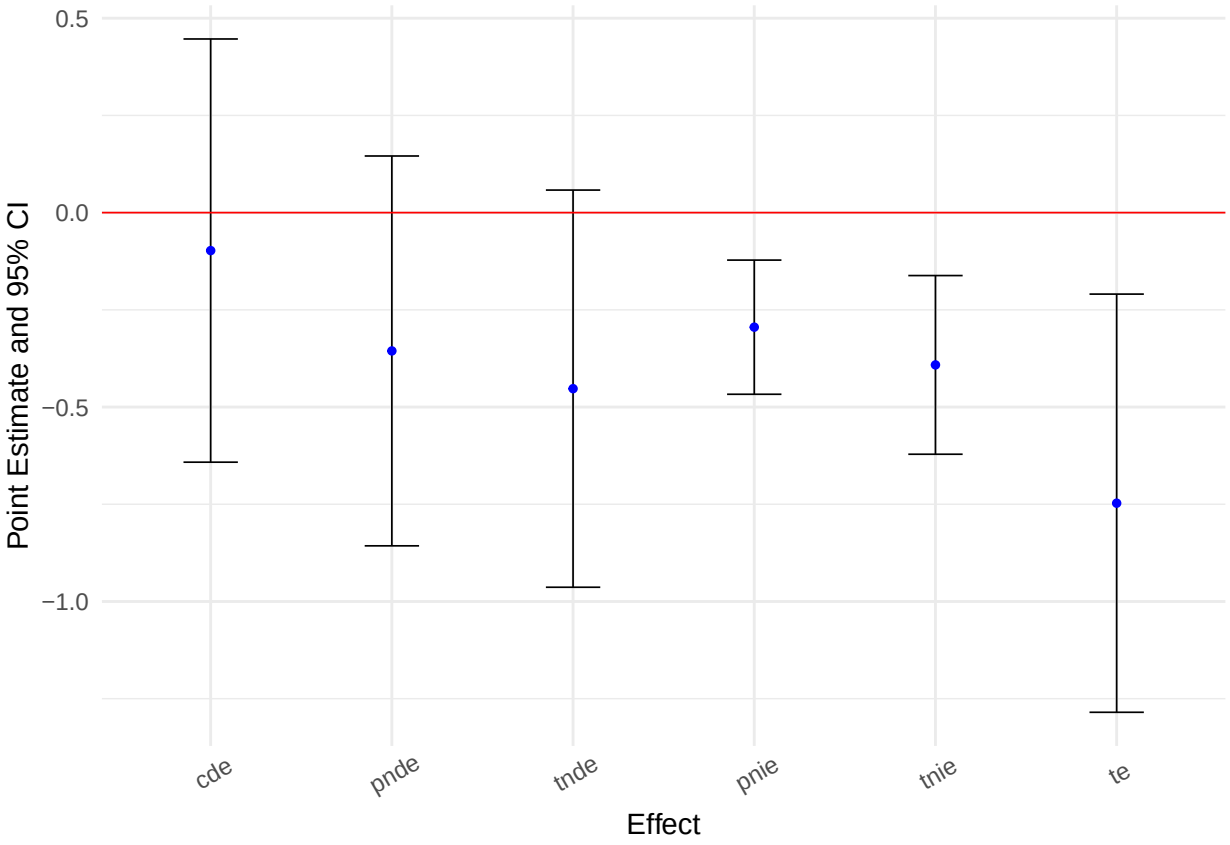
```r
  basec = c("female", "approxage", "protein_c"),
  EMint = TRUE,  # allow exposure-mediator interaction
  mreg = list(type = "linear"),
  yreg = "linear",
  estimation = "paramfunc",
  inference = "delta",
  full = FALSE,
  astar = 0, # reference value
  a = 1,
  mval = list(1)
)

summary(med_result)
```

```
## Causal Mediation Analysis
##
## # Outcome regression:
##
## Call:
## glm(formula = cognitive_raw ~ ln_mn_c + birthlength_c + ln_mn_c *
##     birthlength_c + female + approxage + protein_c, family = gaussian(),
##     data = getCall(x$reg.output$yreg)$data, weights = getCall(x$reg.output$yreg)$weights)
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           56.15984    1.36760  41.064  < 2e-16 ***
## ln_mn_c               -0.37497    0.25480  -1.472  0.14176
## birthlength_c          0.84189    0.12446   6.764 3.81e-11 ***
## female1               -0.08317    0.48416  -0.172  0.86367
## approxage              0.12095    0.05800   2.085  0.03756 *
## protein_c              0.15433    0.05357   2.881  0.00414 **
## ln_mn_c:birthlength_c  0.27738    0.12058   2.300  0.02185 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 28.95798)
##
##     Null deviance: 16763  on 499  degrees of freedom
## Residual deviance: 14276  on 493  degrees of freedom
## AIC: 3110.8
##
## Number of Fisher Scoring iterations: 2
##
##
## # Mediator regressions:
##
## Call:
## glm(formula = birthlength_c ~ ln_mn_c + female + approxage +
##     protein_c, family = gaussian(), data = getCall(x$reg.output$mreg[[1L]])$data,
##     weights = getCall(x$reg.output$mreg[[1L]])$weights)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  0.377127   0.495002   0.762 0.446500
## ln_mn_c     -0.349862   0.090862  -3.850 0.000133 ***
## female1     -0.275253   0.174827  -1.574 0.116028
## approxage   -0.006579   0.021002  -0.313 0.754205
## protein_c    0.098298   0.018891   5.203 2.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 3.79841)
##
##     Null deviance: 2048.0  on 499  degrees of freedom
## Residual deviance: 1880.2  on 495  degrees of freedom
## AIC: 2093.2
##
## Number of Fisher Scoring iterations: 2
##
##
## # Effect decomposition on the mean difference scale via the regression-based approach
##
## Closed-form parameter function estimation with
##  delta method standard errors, confidence intervals and p-values
##
##      Estimate Std.error  95% CIL 95% CIU    P.val
## cde  -0.09759   0.27764 -0.64176   0.447 0.725209
## pnde -0.35562   0.25576 -0.85691   0.146 0.164393
## tnde -0.45267   0.26057 -0.96337   0.058 0.082347 .
## pnie -0.29455   0.08802 -0.46706  -0.122 0.000819 ***
## tnie -0.39159   0.11721 -0.62132  -0.162 0.000835 ***
## te   -0.74721   0.27436 -1.28494  -0.209 0.006459 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (cde: controlled direct effect; pnde: pure natural direct effect; tnde: total natural direct effect;
##
## Relevant variable values:
## $a
## [1] 1
##
## $astar
## [1] 0
##
## $mval
## $mval[[1]]
## [1] 1
##
##
## $basecval
## $basecval[[1]]
## [1] 0.5
##
## $basecval[[2]]
## [1] 23.00725
##
## $basecval[[3]]
```

```
## [1] -0.1868748
```

```
# visualization
ggcmest(med_result) +
ggplot2::theme(axis.text.x = ggplot2::element_text(angle = 30, vjust = 0.8))
```



- Pure natural direct effect (PNDE): -0.356 (95%CI -0.857, 0.146)
- Total natural direct effect (TNDE): -0.453 (95%CI -0.963, 0.058)
- Pure natural indirect effect (PNIE): -0.295 (95%CI -0.467, -0.122)
- Total natural indirect effect (TNIE): -0.392 (95%CI -0.621, -0.162)
- Total effect: -0.747 (95%CI -1.285, -0.209)

2. Interpretation and assumptions

- Natural direct effect (NDE): The NDE compares the outcome Y under the treatment $A = 1$ to the outcome under no treatment $A = 0$, keeping the mediator M fixed at its natural level when $A = 0$. In this analysis, the NDE represents how manganese exposure directly impacts cognitive scores, independent of its effect through birth length. When assuming no exposure-mediator interaction, NDE was negative (-0.356) but it was not statistically significant (p-value = 0.164). When accounting for exposure-mediator interaction, NDE was also negative (-0.453) but it was not statistically significant (p-value = 0.0823)
- Natural indirect effect (NIE): The NIE compares the outcome under $A = 1$ where the mediator M takes its natural value under $A = 1$, to the outcome under $A = 1$ where the mediator is set to its natural value under $A = 0$. In this analysis, the NIE quantifies the mediated pathway, i.e., the effect of manganese on

cognitive scores that operates through birth length. When assuming no exposure-mediator interaction, this effect was significant (p-value <0.01) with an estimate of -0.295, indicating that shorter birth lengths associated with higher manganese exposure negatively impact cognitive outcomes. This trend persisted when accounting for exposure-mediator interaction (estimate: -0.392, p-value <0.01)
- Total effect: The TE is the overall effect of changing the treatment from $A = 0$ to $A = 1$, which is decomposable into NDE + NIE. In this analysis, the TE was significant (p-value <0.01) with an estimate of -0.747, confirming that increased manganese exposure reduces cognitive scores

- In summary, the above results suggests that the mediator explains a substantial part of the observed total effect

The following identifiability assumptions are necessary for the causal interpretation:

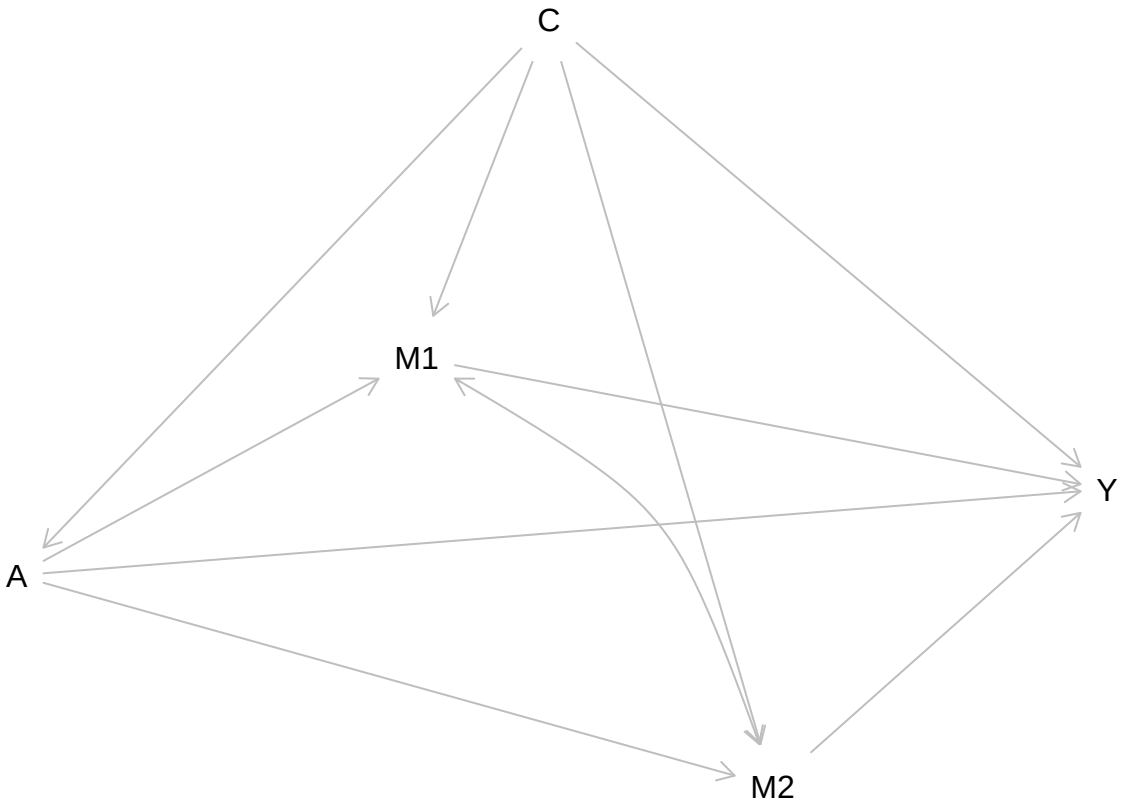
- $Y_{am} \perp A \mid C$: No unmeasured confounders affect the relationship between exposure A and outcome Y after adjusting for covariates C
- $Y_{am} \perp M \mid C, A$: No unmeasured confounders influence the relationship between the mediator M and outcome Y, conditional on exposure and covariates
- $M_a \perp A \mid C$: No unmeasured confounders affect the relationship between exposure A and mediator M after adjusting for covariates
- $Y_{am} \perp M_{a*} \mid C$: No effect of exposure A that confounds the mediator M and outcome Y relationship

3. Other measures of fetal growth

(a) Causal DAG
Exposure: Manganese levels in the mother (A)
Mediator: Birth length (M_1), additional fetal growth measurement such as birthweight (M_2)
Outcome: Cognitive score (Y)
Covariates: Potential confounders such as maternal age, sex of the child, and protein intake (C)

Below is the DAG with the additional measurements of fetal growth:

```
# DAG
g = dagitty('dag {
A [pos="-1.675,-0.260"]
C [pos="-0.451,-1.098"]
M1 [pos="-0.755,-0.588"]
M2 [pos="0.064,0.058"]
Y [pos="0.833,-0.390"]
A -> M1
A -> M2
A -> Y
C -> A
C -> M1
C -> M2
C -> Y
M1 -> Y
M1 <-> M2
M2 -> Y
}')

plot(g)
```



(b) Adjustment of birthweight and identifiability assumption Adjustment of birthweight (M_2) could lead to a violation of the fourth identifiability assumption if it induces collider bias or blocks a portion of the causal pathway.
In this context, M_2 is likely to be a collider as birthweight could be affected by maternal age, sex of the child, and protein intake. Birthweight might also mediate the relationship independently or jointly with birth length. By adjusting for birthweight, indirect effect might be partially removed, violating the decomposition of total effects into direct and indirect effects.