Simulation code - 8월 6일 Version

1. $\widehat{\mu}_0$과 $\widehat{\mu}_1$의 추정량 변경해보고 IPW ATT Variance estimator Hardcoding해 구한 값이랑 함수 쓴 결과 다시 비교

Ver 04
$$\hat{\mu}_0 = \log 1.5 \times P(B \mid E=1) + \log 2 \cdot E[C \mid E=1]$$
$$\hat{\mu}_1 = \hat{\mu}_0 + \log 2$$

Ver 05
$\hat{\mu}_1$ = mean(result_model$fitted.values[data$E==1])
$\hat{\mu}_0$ = mean(result_model$fitted.values[data$E==0])

Ver 08
$$\hat{\mu}_1 = \frac{\sum \hat{\omega}_i A_i Y_i}{\sum \hat{\omega}_i A_i}$$
$$\hat{\mu}_0 = \frac{\sum \hat{\omega}_i (1-A_i) Y_i}{\sum \hat{\omega}_i (1-A_i)}$$
($\hat{\omega}_i$는 weight 의미)

[Result]

	Estimator		Result	
Scenario 1	mu0_hat	Ver04	function value	0.1152347
	mu1_hat	Ver04	package value	0.08249028
Scenario 2	mu0_hat	Ver04	function value	0.09211579
	mu1_hat	Ver05	package value	0.08249028

: 생각한 조합을 모두 고려해보았을 때, Scenario 2 조합이 실제 값과의 차이가 가장 적다.

2. Doubly robust estimator

1) Doubly robust ATE estimator

$$\begin{aligned}
\hat{\tau}_{\mathrm{dr}} &= \hat{\mu}_{1,\mathrm{dr}} - \hat{\mu}_{0,\mathrm{dr}} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{Z_i Y_i}{\hat{e}(X_i)} - \frac{Z_i - \hat{e}(X_i)}{\hat{e}(X_i)}\hat{m}_1(X_i)\right\} - \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{(1-Z_i)Y_i}{1-\hat{e}(X_i)} + \frac{Z_i - \hat{e}(X_i)}{1-\hat{e}(X_i)}\hat{m}_0(X_i)\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\hat{m}_1(X_i) + \frac{Z_i\{Y_i - \hat{m}_1(X_i)\}}{\hat{e}(X_i)}\right] - \frac{1}{N}\sum_{i=1}^{N}\left[\hat{m}_0(X_i) + \frac{(1-Z_i)\{Y_i - \hat{m}_0(X_i)\}}{1-\hat{e}(X_i)}\right]
\end{aligned}$$

: 여기서 $\widehat{m}_1(X_i)$은 E[Y|E=1, X]을, $\widehat{m}_0(X_i)$은 E[Y|E=0, X]을 의미한다. --> Outcome regression part (아래 첨자가 E=e인 sub-population 의미)

2) Doubly robust ATT estimator

$$\hat{\tau}_{\mathrm{dr}}^{\mathrm{ATT}} = \sum_{i=1}^{N} \left[Y_i Z_i - \frac{Y_i(1 - Z_i)\hat{e}_i + \hat{m}_0(\mathbf{X}_i)(Z_i - \hat{e}_i)}{1 - \hat{e}_i} \right] \Big/ N_1,$$

R code)

```
##################################################
##### Doubly robust estimator function #####
DR_estimator<-function(estimate,data,var_treat,var_y,cov){

  PS_df<-weight_make(var_treat,cov,estimate,data)
  data$ps<-PS_df$ps

  myformula<-as.formula(sprintf("%s~.",var_y))

  ind_mu0<-which(data[,var_treat]==0)
  mu0_df<-data[ind_mu1,c(var_y,cov)]

  out_mu0<-lm(formula=myformula,data=mu0_df)
  mu0_X<-coef(out_mu0)%*%t(cbind(1,data[,cov]))

  if(estimate=='ATE'){
    ind_mu1<-which(data[,var_treat]==1)
    mu1_df<-data[ind_mu1,c(var_y,cov)]
    out_mu1<-lm(formula=myformula,data=mu1_df)
    mu1_X<-coef(out_mu1)%*%t(cbind(1,data[,cov]))

    mu1_dr<-mean(data[,var_treat]*data[,var_y]/data$ps - ((data[,var_treat]-data$ps)/data$ps)*(coef(out_mu1)%*%t(cbind(1,data[,cov]))))
    mu0_dr<-mean(mu0_X+((1-data[,var_treat])*(data[,var_y]-mu0_X))/(1-data$ps))

    result<-mu1_dr-mu0_dr
    }

  else if(estimate=='ATT'){
    ind_mu1<-which(data[,var_treat]==1)

    result<-sum(data[,var_y]*data[,var_treat]-((data[,var_y]*(1-data[,var_treat])*data$ps)+(mu0_X)*(data[,var_treat]-data$ps))/(1-data$ps))/length(ind_mu1)
  }

  return(result)
}
```

Checking)

```
######### check #####################################################
E<-E_sample[,1]
B<-B_sample[,1]
#U<-U_sample[,1]
C<-C_sample[,1]
Y<-Y_sample[,1]
data<-data.frame("E"=E,"B"=B,"C"=C,"Y"=Y)
cov<-c("B","C")
cov_type<-c("binary","continuous")
#mydata<-data_export("E",cov,data)
#head(mydata)

DR_ATE<-DR_estimator("ATE",data,"E","Y",cov)
DR_ATT<-DR_estimator("ATT",data,"E","Y",cov)
```

[Result]

	# of obs	Doubly robust estimator	IPW estimator	True value
ATT	1000	0.7174175	0.7482638	0.6931472
ATE	1000	0.7394487	0.7482638	0.6931472

Question) Outcome model을 “lm” 사용해도 무방한가?

----

3. Naive variance estimator of DR estimator

참고한 공식

DR estimator: Variance

- Lunceford and Davidian (2004) provides an estimator to approximate the variance of $\hat{\tau}_{\text{dr}}$:

$$s^2_{\text{dr}} = \sum_i (\hat{\tau}_i - \hat{\tau}_{\text{dr}})^2 / N^2, \tag{2}$$

where

$$\hat{\tau}_i = \left[\frac{Z_i Y_i}{\hat{e}_i} - \frac{(Z_i - \hat{e}_i)\hat{m}_1(\mathbf{X}_i)}{\hat{e}_i} \right] - \left[\frac{(1 - Z_i)Y_i}{(1 - \hat{e}_i)} + \frac{(Z_i - \hat{e}_i)\hat{m}_0(\mathbf{X}_i)}{(1 - \hat{e}_i)} \right]$$

R code

```
##########################################################################
####### Naive variance estimator of DR estimator ######
DR_Naive_var_estimator<-function(estimate,data,var_treat,var_y,cov){
  result<-DR_estimator(estimate,data,var_treat,var_y,cov)

  tau_dr<-result$est
  tau_i<-data[,var_treat]*data[,var_y]/data$ps -
    ((data[,var_treat]-data$ps)/data$ps)*(coef(out_mu1)%*%t(cbind(1,data[,cov])))-
    mu0_X+((1-data[,var_treat])*(data[,var_y]-mu0_X))/(1-data$ps)

  se<-sum((tau_i-tau_dr)^2)/(nrow(data)^2)

  return(se)

}
##############################################################
```

Package function 사용

: drgee package의 drgee function 사용

: 행렬의 역행렬 계산 과정에서 error

```
> library(drgee)
> result<-drgee(oformula=formula(Y~E),
+               eformula=formula(E~B+C),
+               olink = 'identity',elink='logit',
+               data=data,estimation.method = 'dr')
Error in solve.default(d.U) :
  Lapack routine dgesv: system is exactly singular: U[7,7] = 0
```