

<Simulation Scenario 구현 결과>

- 12월 13일 Version

<WHAT TO DO>

1) ATE, ATT 추정치가 방법론 별 같은지 확인 (Outcome regression estimator의 경우 동일할 수 있음.)

2) 새로운 성능평가지표 $\frac{\widehat{var}(\hat{\beta})}{var(\hat{\beta})}$ 의 분모는 여러 번의 Replication 통해 얻은 추정치들의 표본 분산으로 값 대입

<Result>

1) N=1000, REPL=1000으로 설정한 뒤, (Exposure ratio는 그대로 0.12를 유지) 방법론 별 ATE, ATT 추정치 계산하여 비교해보니 Outcome regression estimator 외에 다른 방법론에서는 결과가 모두 다름을 확인.

```
> table(ATE_1_result$Outcome_reg_est==ATT_1_result$Outcome_reg_est)
FALSE TRUE
  965   35
> table(ATE_1_result$IPW_est==ATT_1_result$IPW_est)
FALSE
 1000
> table(ATE_1_result$DR_est==ATT_1_result$DR_est)
FALSE
 1000
```

2) SD Ratio " $\frac{\widehat{var}(\hat{\beta})}{var(\hat{\beta})}$ " 분모 값을 다양한 version으로 값을 얻은 뒤 Performance result table 얻어봄.

2)-①. $var(\hat{\beta})$: N=100000, REPL=1000번 해서 각 방법론 별 ATE, ATT 추정치 1000개 얻은 뒤 해당 추정치들의 표본 분산 값 대입

[Result]

```
print(ATE_1_performance)
#               Bias               rMSE Naive_var_coverage Naive_var_SD_Ratio Sandwich_robust_var_coverage Sandwich_var_SD_Ratio
# Outcome_reg -0.002379192 0.009928812             1.000             1077.3445                NA                NA
# IPW          -0.002150330 0.010115352             1.000              859.5440                0.982             161.6638
# DR           -0.002304139 0.009961594             0.938              105.7955                0.980             131.5339

print(ATT_1_performance)
#               Bias               rMSE Naive_var_coverage Naive_var_SD_Ratio Sandwich_robust_var_coverage Sandwich_var_SD_Ratio
# Outcome_reg -0.002379192 0.009928812             1.000             50752.623671                NA                NA
# IPW          -0.002438935 0.009934539             1.000              858.489100                0.995             212.9208
# DR           -0.002399891 0.009940280             0.377               6.792995                1.000             1151.2080
```

2)-②. $var(\hat{\beta})$: N=100000으로 설정한 뒤, 각 방법론 별 ATE, ATT 추정량의 분산 추정량 오직 1번만 구해 해당 값을 대입

[Result]

----- 결과를 보지 못함 -----

<Question>

1) Outcome regression estimator 선택이 고민

: Outcome regression model 적합한 뒤 처치변수에 대응하는 회귀계수를 사용해야 하는지
/ 검색해 찾은 공식을 이용해야 하는지

▶ For ATE, the outcome-regression-based estimator is

$$\hat{\tau}^{ATE} = \sum_{i=1}^N Z_i(Y_i - \hat{\mu}_0(X_i)) + (1 - Z_i)(\hat{\mu}_1(X_i) - Y_i) \Big/ N$$

▶ For ATT, the outcome-regression-based estimator is

$$\hat{\tau}^{ATT} = \sum_{i=1}^N Z_i \{Y_i - \hat{\mu}_0(X_i)\} \Big/ N_1$$

2) Outcome regression ATT Estimator의 Naive variance estimator 공식 재확인 필요
(My formula)

```
a<-(A * Y - A *coef(Outcome regression model))%% data.frame(cbind(1, B, C, 0)))  
var<-sum((a-ATT_outcome_reg_est)^2) / length(A==1)
```

3) Sandwich robust variance of IPW ATT 추정치가 NaN이 나오는 경우가 종종 존재

: 나올 수 있는 경우인지 / 함수 어느 부분에서 나올 수 있는 부분인지