Appendix

ZiqianHe

d

```
mean(Y[A==1])-mean(Y[A==0])
```

```
## [1] -2.566766
```

e

```
dt = data.table(Y=Y,A=A,C=C)

#without boostrap
mean(dt[A==1][C==0]$Y)
```

```
## [1] 114.5484
```

```
mean(dt[A==1][C==1]$Y) #means no data
```

```
## [1] NaN
```

```
mean(dt[A==0][C==0]$Y)
```

```
## [1] 109.7194
```

```
mean(dt[A==0][C==1]$Y)
```

```
## [1] 129.4415
```

```
E_Y1 = mean(dt[A==1][C==0]$Y)*0.8
E_Y0 = mean(dt[A==0][C==0]$Y)*0.8 + mean(dt[A==0][C==1]$Y)*0.2
E_Y1-E_Y0
```

```
## [1] -22.02507
```

```r
#boostrap
dt$interv <- -1 # 1st copy: equal to original one

interv0 <- dt # 2nd copy: treatment set to 0, outcome to missing
interv0$interv <- 0
interv0$A <- 0
interv0$Y <- NA

interv1 <- dt # 3rd copy: treatment set to 1, outcome to missing
interv1$interv <- 1
interv1$A <- 1
interv1$Y <- NA

onesample <- rbind(dt, interv0, interv1) # combining datasets

standardization <- function(data, indices) {
  # create a dataset with 3 copies of each subject
  d <- data[indices, ] # 1st copy: equal to original one`
  d$interv <- -1
  d0 <- d # 2nd copy: treatment set to 0, outcome to missing
  d0$interv <- 0
  d0$A <- 0
  d0$Y <- NA
  d1 <- d # 3rd copy: treatment set to 1, outcome to missing
  d1$interv <- 1
  d1$A <- 1
  d1$Y <- NA
  d.onesample <- rbind(d, d0, d1) # combining datasets

  # linear model to estimate mean outcome conditional on treatment and confounders
  # parameters are estimated using original observations only (interv= -1)
  # parameter estimates are used to predict mean outcome for observations with set
  # treatment (interv=0 and interv=1)
  fit <- glm(
    Y ~ A + as.factor(C),
    data = d.onesample
  )

  d.onesample$predicted_meanY <- predict(fit, d.onesample)

  # estimate mean outcome in each of the groups interv=-1, interv=0, and interv=1
  return(c(
    mean(d.onesample$predicted_meanY[d.onesample$interv == -1]),
    mean(d.onesample$predicted_meanY[d.onesample$interv == 0]),
    mean(d.onesample$predicted_meanY[d.onesample$interv == 1]),
    mean(d.onesample$predicted_meanY[d.onesample$interv == 1]) -
      mean(d.onesample$predicted_meanY[d.onesample$interv == 0])
  ))
}

# bootstrap
results <- boot(data = dt,
                statistic = standardization,
```

```
             R = 5)

# generating confidence intervals
se <- c(sd(results$t[, 1]),
        sd(results$t[, 2]),
        sd(results$t[, 3]),
        sd(results$t[, 4]))
mean <- results$t0
ll <- mean - qnorm(0.975) * se
ul <- mean + qnorm(0.975) * se

bootstrap <-
  data.frame(cbind(
    c(
      "Observed",
      "No Treatment",
      "Treatment",
      "Treatment - No Treatment"
    ),
    mean,
    se,
    ll,
    ul
  ))

bootstrap
```

```
##                             V1             mean                 se                 ll
## 1                     Observed 115.831804247806  1.01744285848755 113.837652888843
## 2                 No Treatment 113.417282662121  1.28581362142713 110.897134273293
## 3                    Treatment  118.24632583349  1.15864358192083 115.975426142007
## 4 Treatment - No Treatment 4.82904317136916 0.530261148976296  3.7897504169748
##                ul
## 1 117.825955606769
## 2 115.937431050949
## 3 120.517225524974
## 4 5.86833592576353
```