Matching Estimators in R

Shenyang Guo and Peter Sun

September, 2021

# Load Packages

library(haven)  
library(lmtest)  
library(tidyverse)  
library(broom)  
library(Matching)  
library(tidyr)  
library(flextable)  
select <- dplyr::select

# Load Data

d <- read\_dta("data/cds\_pcss97.dta") %>%  
 haven::zap\_formats() %>%  
 sjlabelled::remove\_all\_labels() %>%  
 as\_tibble()

# Breusch-Pagan Test for Heteroskedasticity

m0 <- lm(pcss97 ~ kuse + male + black + age97 + pcged97 + mratio96 + pcg\_adc, data = d)  
get\_bptest <- function(data, lm.model, var) {  
 b <- lmtest::bptest(lm.model, as.formula(paste0("~", var)),   
 data = data, studentize = F)  
 return(tibble(variable = var, statistic = b$statistic,   
 df = b$parameter, p.value = b$p.value))  
}  
map\_dfr(c("kuse", "male", "black", "age97", "pcged97", "mratio96", "pcg\_adc"),   
 get\_bptest, data = d, lm.model = m0) %>%  
 flextable() %>%  
 flextable::set\_table\_properties(width = 1, layout = "autofit")

| variable | statistic | df | p.value |
| --- | --- | --- | --- |
| kuse | 0.009620929 | 1 | 0.9218637379735 |
| male | 3.070854377 | 1 | 0.0797075706437 |
| black | 1.434498331 | 1 | 0.2310317091301 |
| age97 | 25.269838865 | 1 | 0.0000004984423 |
| pcged97 | 0.189052730 | 1 | 0.6637061942448 |
| mratio96 | 0.618238384 | 1 | 0.4317026677797 |
| pcg\_adc | 0.291050245 | 1 | 0.5895482933337 |

# Matching Estimators

## Define Outcome (Y), Treatment Index (Tr), and Variables to Match On (X)

Y <- d$pcss97  
Tr <- d$kuse  
X <- select(d, male, black, age97, pcged97, mratio96, pcg\_adc)

## Get Estimators Individually

# Sample Average Treatment Effect (SATE)  
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,  
 estimand = "ATE", sample = T))  
  
# Population Average Treatment Effect (PATE)  
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,  
 estimand = "ATE", sample = F))  
  
# Sample average treatment effect for the treated (SATT)  
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,  
 estimand = "ATT", sample = T))  
  
# Population average treatment effect for the treated (PATT)  
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,  
 estimand = "ATT", sample = F))  
  
# Sample average treatment effect for the controls (SATC)  
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,  
 estimand = "ATC", sample = T))  
  
# Population average treatment effect for the controls (PATC)  
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,  
 estimand = "ATC", sample = F))

## Get All Estimators

get\_match <- function(estimand, sample) {  
 m <- Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = TRUE, Var.calc = 4,  
 estimand = estimand, sample = sample)  
 return(list(  
 est = m$est[,1],  
 se = m$se,  
 t.stat = m$est[,1]/m$se,  
 p = (1 - pnorm(abs(m$est[,1]/m$se))) \* 2  
 ))  
}  
tribble(  
 ~estimator, ~estimand, ~sample,  
 "SATE", "ATE", T,  
 "PATE", "ATE", F,  
 "SATT", "ATT", T,  
 "PATT", "ATT", F,  
 "SATC", "ATC", T,  
 "PATC", "ATC", F  
) %>%  
 rowwise() %>%  
 mutate(match = list(get\_match(estimand, sample))) %>%  
 tidyr::unnest\_wider(match) %>%  
 select(-estimand, -sample) %>%  
 flextable::flextable() %>%  
 flextable::set\_table\_properties(width = 1, layout = "autofit")

| estimator | est | se | t.stat | p |
| --- | --- | --- | --- | --- |
| SATE | -4.703773 | 1.769696 | -2.657956 | 0.007861620 |
| PATE | -4.703773 | 1.765187 | -2.664746 | 0.007704656 |
| SATT | -5.229651 | 1.781217 | -2.935999 | 0.003324759 |
| PATT | -5.229651 | 1.720590 | -3.039451 | 0.002370097 |
| SATC | -4.467254 | 2.133536 | -2.093827 | 0.036275403 |
| PATC | -4.467254 | 2.135647 | -2.091757 | 0.036460232 |