exercise5

Xiang Li

2024/3/17

Exercise 1

```
data_df = data.frame(X = c(0, 0, 0, 0, 1, 1, 1, 1), C = c(0, 0, 1, 1, 0, 0, 1, 1),
    Y = c(0, 1, 0, 1, 0, 1, 0, 1), n = c(80, 20, 20, 10, 80, 20, 80, 40))
```

a

```
PX1cC0 = sum(data_df[(data_df$X == 1) & (data_df$C == 0), "n"])/sum(data_df[data_df$C ==
    0, "n"])
print(PX1cC0)
```

```
## [1] 0.5
```

```
PX1cC1 = sum(data_df[(data_df$X == 1) & (data_df$C == 1), "n"])/sum(data_df[data_df$C ==
    1, "n"])
print(PX1cC1)
```

```
## [1] 0.8
```

These probabilities are propensity scores of X=1.

b

```
PX0cC0 = 1 - PX1cC0
PX0cC1 = 1 - PX1cC1
WX1cC0 = 1/PX1cC0
print(WX1cC0)
```

```
## [1] 2
```

```
WX1cC1 = 1/PX1cC1
print(WX1cC1)
```

```
## [1] 1.25
```

```r
WX0cC0 = 1/PX0cC0
print(WX0cC0)
```

```
## [1] 2
```

```r
WX0cC1 = 1/PX0cC1
print(WX0cC1)
```

```
## [1] 5
```

c

```r
data_df$W = 0
data_df[(data_df$X == 1) & (data_df$C == 0), "W"] = WX1cC0
data_df[(data_df$X == 1) & (data_df$C == 1), "W"] = WX1cC1
data_df[(data_df$X == 0) & (data_df$C == 0), "W"] = WX0cC0
data_df[(data_df$X == 0) & (data_df$C == 1), "W"] = WX0cC1
data_df$YxnxW = data_df$Y * data_df$n * data_df$W
data_df$nxW = data_df$n * data_df$W
EY1 = sum(data_df[data_df$X == 1, "YxnxW"])/sum(data_df[data_df$X == 1, "nxW"])
print(EY1)
```

```
## [1] 0.2571429
```

```r
EY0 = sum(data_df[data_df$X == 0, "YxnxW"])/sum(data_df[data_df$X == 0, "nxW"])
print(EY0)
```

```
## [1] 0.2571429
```

```r
ATE = EY1 - EY0
print(ATE)
```

```
## [1] 0
```

Exercise 2

```r
library(cobalt)
```

```
## Warning: package 'cobalt' was built under R version 4.3.2
```

```
##  cobalt (Version 4.5.4, Build Date: 2024-02-26)
```

```r
library(survey)
```

```
## Warning: package 'survey' was built under R version 4.3.2
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```r
library(ggplot2)
library(gridExtra)
library(survey)
```

```r
load("rhc_exercise.RData")
```

a

```r
ps_model = glm(treatment ~ transhx + age + scoma1 + hrt1 + bili1 + wtkilo1 + cat1,
    data = rhc, family = "binomial")
print(summary(ps_model))
```

```
##
## Call:
## glm(formula = treatment ~ transhx + age + scoma1 + hrt1 + bili1 +
##     wtkilo1 + cat1, family = "binomial", data = rhc)
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.9073406  0.1967016  -9.697  < 2e-16 ***
## transhx               0.5426888  0.0875449   6.199 5.68e-10 ***
## age                   0.0032199  0.0017884   1.800  0.07179 .
## scoma1               -0.0019473  0.0011566  -1.684  0.09224 .
## hrt1                  0.0028841  0.0007398   3.899 9.67e-05 ***
## bili1                 0.0267466  0.0061889   4.322 1.55e-05 ***
## wtkilo1               0.0098138  0.0013141   7.468 8.13e-14 ***
## cat1CHF               0.3973320  0.1064659   3.732  0.00019 ***
## cat1COPD             -1.3804068  0.1485325  -9.294  < 2e-16 ***
## cat1Cirrhosis        -0.9184740  0.1743862  -5.267 1.39e-07 ***
## cat1Colon Cancer     -1.1519303  1.0871074  -1.060  0.28931
## cat1Coma             -0.5785267  0.1429332  -4.048 5.18e-05 ***
```

```
## cat1Lung Cancer       -1.2690193  0.4826942  -2.629  0.00856 **
## cat1MOSF w/Malignancy  0.0984301  0.1127698   0.873  0.38275
## cat1MOSF w/Sepsis      0.7993355  0.0727869  10.982  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7621.4  on 5734  degrees of freedom
## Residual deviance: 7049.6  on 5720  degrees of freedom
## AIC: 7079.6
##
## Number of Fisher Scoring iterations: 4
```

Except age and scoma1, other variables are strongly associated with treatment.

b

```
rhc$ps1 = predict(ps_model, newdata = rhc, type = "response")
rhc$ipw1 = 1/rhc$ps1
rhc$ipw1[rhc$treatment == "no RHC"] = 1/(1 - rhc$ps1[rhc$treatment == "no RHC"])
```

c

```
vars1 = c("transhx", "age", "scoma1", "hrt1", "bili1", "wtkilo1", "cat1", "surv2md1",
    "aps1")
covariates = rhc[, vars1]
table1 = bal.tab(covariates, treat = rhc$treatment, weights = rhc$ipw1, method = "weighting",
    un = TRUE)
```

```
## Note: `s.d.denom` not specified; assuming "pooled".
```

```
print(table1)
```

```
## Balance Measures
##                            Type Diff.Un Diff.Adj
## transhx                  Binary  0.0554   0.0007
## age                     Contin. -0.0614   0.0005
## scoma1                  Contin. -0.1098   0.0227
## hrt1                    Contin.  0.1469  -0.0164
## bili1                   Contin.  0.1446  -0.0075
## wtkilo1                 Contin.  0.2256  -0.0012
## cat1_ARF                 Binary -0.0290  -0.0040
## cat1_CHF                 Binary  0.0261   0.0000
## cat1_COPD                Binary -0.0858   0.0002
## cat1_Cirrhosis           Binary -0.0268   0.0001
## cat1_Colon Cancer        Binary -0.0012   0.0000
## cat1_Coma                Binary -0.0525   0.0057
```

```
## cat1_Lung Cancer        Binary -0.0073  -0.0003
## cat1_MOSF w/Malignancy  Binary  0.0045  -0.0008
## cat1_MOSF w/Sepsis      Binary  0.1721  -0.0010
## surv2md1               Contin. -0.1985  -0.2997
## aps1                   Contin.  0.5014   0.4165
##
## Effective sample sizes
##             no RHC     RHC
## Unadjusted 3551.   2184.
## Adjusted   3297.72 1738.58
```

```
# love.plot(covariates, treat = rhc$treatment, weights = rhc$ipw1, method =
# 'weighting', binary = 'std', threshold = .1)
```

After propensity score weight adjustment by transhx, age, scoma1, hrt1, bili1, wtkilo1, cat1, the treatment is still imbalanced on surv2md1 and aps1.

d

```
ps_model1 = glm(treatment ~ transhx + age + scoma1 + hrt1 + bili1 + wtkilo1 + cat1 +
    surv2md1 + aps1, data = rhc, family = "binomial")
print(summary(ps_model1))
```

```
##
## Call:
## glm(formula = treatment ~ transhx + age + scoma1 + hrt1 + bili1 +
##     wtkilo1 + cat1 + surv2md1 + aps1, family = "binomial", data = rhc)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -1.0372749  0.3362492  -3.085  0.00204 **
## transhx                0.6860904  0.0904192   7.588 3.25e-14 ***
## age                   -0.0051040  0.0020279  -2.517  0.01184 *
## scoma1                -0.0068209  0.0013238  -5.153 2.57e-07 ***
## hrt1                   0.0014886  0.0007586   1.962  0.04973 *
## bili1                 -0.0002230  0.0063827  -0.035  0.97213
## wtkilo1                0.0109611  0.0013536   8.098 5.60e-16 ***
## cat1CHF                0.9062462  0.1146630   7.904 2.71e-15 ***
## cat1COPD              -1.0939852  0.1519484  -7.200 6.03e-13 ***
## cat1Cirrhosis         -0.9407151  0.1779204  -5.287 1.24e-07 ***
## cat1Colon Cancer      -1.0907678  1.1049168  -0.987  0.32355
## cat1Coma              -0.7041024  0.1532261  -4.595 4.32e-06 ***
## cat1Lung Cancer       -1.1897262  0.4922679  -2.417  0.01566 *
## cat1MOSF w/Malignancy -0.6373686  0.1340213  -4.756 1.98e-06 ***
## cat1MOSF w/Sepsis      0.6492980  0.0754932   8.601  < 2e-16 ***
## surv2md1              -1.8602550  0.2608313  -7.132 9.89e-13 ***
## aps1                   0.0177368  0.0019704   9.002  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

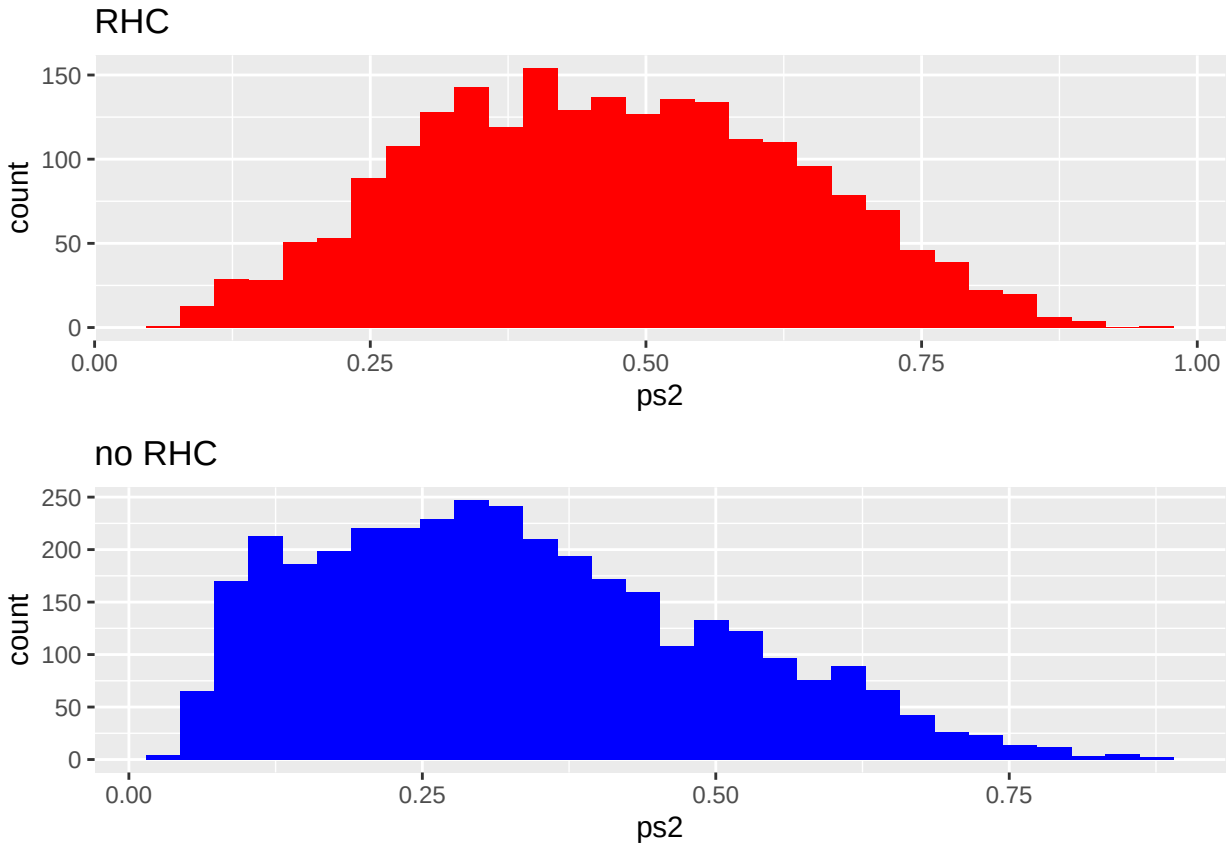
```
##
##     Null deviance: 7621.4  on 5734  degrees of freedom
## Residual deviance: 6779.1  on 5718  degrees of freedom
## AIC: 6813.1
##
## Number of Fisher Scoring iterations: 4
```

```
rhc$ps2 = predict(ps_model1, newdata = rhc, type = "response")
print(summary(rhc$ps2))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.03975 0.24215 0.36314 0.38082 0.51358 0.96591
```

```
p1 = ggplot(data = rhc[rhc$treatment == "RHC", ], mapping = aes(x = ps2)) + geom_histogram(fill = "red"
    labs(title = "RHC")
p2 = ggplot(data = rhc[rhc$treatment == "no RHC", ], mapping = aes(x = ps2)) + geom_histogram(fill = "bl
    labs(title = "no RHC")
grid.arrange(p1, p2, nrow = 2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The minimal ps2 is 0.04 and the maximal ps2 is 0.97, which means no individual will get a propensity score 1 or 0. The histograms show overlap between two treatment groups. Positivity assumption holds.

e

```
rhc$ipw2 = (rhc$treatment == "RHC") * 1/rhc$ps2 + (rhc$treatment == "no RHC") * 1/(1 -
    rhc$ps2)
table2 = bal.tab(covariates, treat = rhc$treatment, weights = rhc$ipw2, method = "weighting",
    un = TRUE)
```

```
## Note: `s.d.denom` not specified; assuming "pooled".
```

```
print(table2)
```

```
## Balance Measures
##                             Type Diff.Un Diff.Adj
## transhx                   Binary  0.0554   0.0018
## age                      Contin. -0.0614  -0.0023
## scoma1                   Contin. -0.1098  -0.0034
## hrt1                     Contin.  0.1469   0.0150
## bili1                    Contin.  0.1446   0.0060
## wtkilo1                  Contin.  0.2256   0.0045
## cat1_ARF                  Binary -0.0290   0.0061
## cat1_CHF                  Binary  0.0261   0.0034
## cat1_COPD                 Binary -0.0858  -0.0088
## cat1_Cirrhosis            Binary -0.0268  -0.0022
## cat1_Colon Cancer         Binary -0.0012   0.0001
## cat1_Coma                 Binary -0.0525  -0.0055
## cat1_Lung Cancer          Binary -0.0073  -0.0014
## cat1_MOSF w/Malignancy    Binary  0.0045   0.0036
## cat1_MOSF w/Sepsis        Binary  0.1721   0.0047
## surv2md1                 Contin. -0.1985  -0.0187
## aps1                     Contin.  0.5014   0.0235
##
## Effective sample sizes
##              no RHC     RHC
## Unadjusted 3551.   2184.
## Adjusted   3116.93 1680.32
```

The distribution of all covariates become balanced after reweighting by ps2.

f

```
print(table(rhc$death30))
```

```
##
## alive death
##  3817  1918
```

```
rhc$death30 = as.numeric(rhc$death30) - 1
print(table(rhc$death30))
```

```
##
##    0    1
## 3817 1918
```

g

```
EY1 = weighted.mean(x = rhc[rhc$treatment == "RHC", "death30"], w = rhc[rhc$treatment ==
    "RHC", "ipw2"])
EY0 = weighted.mean(x = rhc[rhc$treatment == "no RHC", "death30"], w = rhc[rhc$treatment ==
    "no RHC", "ipw2"])
ATE = EY1 - EY0
print(c(EY1 = EY1, EY0 = EY0, ATE = ATE))
```

```
##        EY1        EY0        ATE
## 0.35799263 0.32073365 0.03725898
```

h

```
d.w = svydesign(~1, weights = rhc$ipw2, data = rhc)
w_reg = svyglm(death30 ~ treatment, design = d.w)
print(summary(w_reg))
```

```
##
## Call:
## svyglm(formula = death30 ~ treatment, design = d.w)
##
## Survey design:
## svydesign(~1, weights = rhc$ipw2, data = rhc)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.32073    0.00851  37.689  < 2e-16 ***
## treatmentRHC  0.03726    0.01432   2.602  0.00929 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2238068)
##
## Number of Fisher Scoring iterations: 2
```

```
print(confint(w_reg))
```

```
##                    2.5 %    97.5 %
## (Intercept)  0.304050792 0.3374165
## treatmentRHC 0.009187664 0.0653303
```

The 95% confidence interval of ATE is (0.009, 0.065)

i

```r
summary(rhc$ipw2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.035   1.340   1.643   1.985   2.240  15.452
```

The individual with very large weights can have big effect on the ATE. If they are outliers, the final results will be biased. If weights are very large, they are usually truncated.

j

```r
ipw2_99th = quantile(rhc$ipw2, probs = 0.99)
rhc$ipw2_trunc = ifelse(rhc$ipw2 > ipw2_99th, ipw2_99th, rhc$ipw2)
EY1 = weighted.mean(x = rhc[rhc$treatment == "RHC", "death30"], w = rhc[rhc$treatment ==
    "RHC", "ipw2_trunc"])
EY0 = weighted.mean(x = rhc[rhc$treatment == "no RHC", "death30"], w = rhc[rhc$treatment ==
    "no RHC", "ipw2_trunc"])
ATE = EY1 - EY0
print(c(EY1 = EY1, EY0 = EY0, ATE = ATE))
```

```
##        EY1        EY0        ATE
## 0.35962584 0.32077150 0.03885434
```

ATE becomes a little bit larger than weighting by ipw2 but still very similar.

Exercise 3

```r
load("rhc_exercise.RData")
```

```r
library(MatchIt)
```

```
##
## Attaching package: 'MatchIt'
```

```
## The following object is masked from 'package:cobalt':
##
##     lalonde
```

a

```r
match_out1 = matchit(treatment ~ transhx + age + scoma1 + hrt1 + bili1 + wtkilo1 +
    cat1 + surv2md1 + aps1, data = rhc, method = "nearest", distance = "glm")
print(summary(match_out1))
```

```
##
## Call:
## matchit(formula = treatment ~ transhx + age + scoma1 + hrt1 +
##     bili1 + wtkilo1 + cat1 + surv2md1 + aps1, data = rhc, method = "nearest",
##     distance = "glm")
##
## Summary of Balance for All Data:
##                        Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                      0.4658        0.3286          0.8097     1.0208
## transhx                       0.1497        0.0943          0.1552          .
## age                          60.7498       61.7609         -0.0647     0.8175
## scoma1                       18.9734       22.2532         -0.1160     0.8116
## hrt1                        118.9281      112.8730          0.1460     1.0260
## bili1                         2.7057        1.9973          0.1329     1.4504
## wtkilo1                      78.2336       73.2753          0.2298     0.9297
## cat1ARF                       0.4162        0.4452         -0.0589          .
## cat1CHF                       0.0957        0.0696          0.0889          .
## cat1COPD                      0.0266        0.1124         -0.5337          .
## cat1Cirrhosis                 0.0224        0.0493         -0.1813          .
## cat1Colon Cancer              0.0005        0.0017         -0.0576          .
## cat1Coma                      0.0435        0.0960         -0.2575          .
## cat1Lung Cancer               0.0023        0.0096         -0.1524          .
## cat1MOSF w/Malignancy         0.0723        0.0679          0.0173          .
## cat1MOSF w/Sepsis             0.3205        0.1484          0.3688          .
## surv2md1                      0.5685        0.6072         -0.1954     1.0663
## aps1                         60.7390       50.9335          0.4837     1.1609
##                        eCDF Mean eCDF Max
## distance                  0.2181   0.3183
## transhx                   0.0554   0.0554
## age                       0.0285   0.0703
## scoma1                    0.0346   0.0683
## hrt1                      0.0344   0.0782
## bili1                     0.0244   0.1119
## wtkilo1                   0.0535   0.1252
## cat1ARF                   0.0290   0.0290
## cat1CHF                   0.0261   0.0261
## cat1COPD                  0.0858   0.0858
## cat1Cirrhosis             0.0268   0.0268
## cat1Colon Cancer          0.0012   0.0012
## cat1Coma                  0.0525   0.0525
## cat1Lung Cancer           0.0073   0.0073
## cat1MOSF w/Malignancy     0.0045   0.0045
## cat1MOSF w/Sepsis         0.1721   0.1721
## surv2md1                  0.0475   0.0957
## aps1                      0.0797   0.2127
##
## Summary of Balance for Matched Data:
##                        Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                      0.4658        0.4209          0.2650     1.4091
## transhx                       0.1497        0.1241          0.0719          .
## age                          60.7498       60.8612         -0.0071     0.7863
## scoma1                       18.9734       18.7161          0.0091     0.9883
## hrt1                        118.9281      117.8668          0.0256     0.9871
## bili1                         2.7057        2.3190          0.0726     1.1183
```

```
## wtkilo1                     78.2336       75.8427          0.1108     0.8740
## cat1ARF                      0.4162        0.4858         -0.1412          .
## cat1CHF                      0.0957        0.1067         -0.0374          .
## cat1COPD                     0.0266        0.0201          0.0399          .
## cat1Cirrhosis                0.0224        0.0270         -0.0309          .
## cat1Colon Cancer             0.0005        0.0000          0.0214          .
## cat1Coma                     0.0435        0.0467         -0.0157          .
## cat1Lung Cancer              0.0023        0.0014          0.0192          .
## cat1MOSF w/Malignancy        0.0723        0.0783         -0.0230          .
## cat1MOSF w/Sepsis            0.3205        0.2340          0.1854          .
## surv2md1                     0.5685        0.5867         -0.0921     0.9837
## aps1                        60.7390       56.9679          0.1860     1.1814
##                         eCDF Mean eCDF Max Std. Pair Dist.
## distance                   0.0593   0.1461          0.2651
## transhx                    0.0256   0.0256          0.6442
## age                        0.0285   0.0641          1.1914
## scoma1                     0.0074   0.0247          0.9372
## hrt1                       0.0089   0.0266          1.0963
## bili1                      0.0144   0.0943          0.5661
## wtkilo1                    0.0297   0.0723          1.0410
## cat1ARF                    0.0696   0.0696          0.8880
## cat1CHF                    0.0110   0.0110          0.6257
## cat1COPD                   0.0064   0.0064          0.2050
## cat1Cirrhosis              0.0046   0.0046          0.3154
## cat1Colon Cancer           0.0005   0.0005          0.0214
## cat1Coma                   0.0032   0.0032          0.4108
## cat1Lung Cancer            0.0009   0.0009          0.0766
## cat1MOSF w/Malignancy      0.0060   0.0060          0.5355
## cat1MOSF w/Sepsis          0.0865   0.0865          0.6387
## surv2md1                   0.0228   0.0582          1.1004
## aps1                       0.0320   0.0966          0.9154
##
## Sample Sizes:
##           Control Treated
## All          3551    2184
## Matched      2184    2184
## Unmatched    1367       0
## Discarded       0       0
```

ATT is being targeted. 2184 treated individuals and 2184 control individuals are matched. After matching, there is still imbalance in some covariates. For example, the standard mean difference of wtkilo1 on matched data is $0.1108 > 0.1$.

b

```
match_out2 = matchit(treatment ~ transhx + age + scoma1 + hrt1 + bili1 + wtkilo1 +
    cat1 + surv2md1 + aps1, data = rhc, method = "nearest", distance = "glm", caliper = 0.1)
print(summary(match_out2))
```

```
##
## Call:
```

```
## matchit(formula = treatment ~ transhx + age + scoma1 + hrt1 +
##     bili1 + wtkilo1 + cat1 + surv2md1 + aps1, data = rhc, method = "nearest",
##     distance = "glm", caliper = 0.1)
##
## Summary of Balance for All Data:
##                        Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                      0.4658        0.3286          0.8097     1.0208
## transhx                       0.1497        0.0943          0.1552          .
## age                          60.7498       61.7609         -0.0647     0.8175
## scoma1                       18.9734       22.2532         -0.1160     0.8116
## hrt1                        118.9281      112.8730          0.1460     1.0260
## bili1                         2.7057        1.9973          0.1329     1.4504
## wtkilo1                      78.2336       73.2753          0.2298     0.9297
## cat1ARF                       0.4162        0.4452         -0.0589          .
## cat1CHF                       0.0957        0.0696          0.0889          .
## cat1COPD                      0.0266        0.1124         -0.5337          .
## cat1Cirrhosis                 0.0224        0.0493         -0.1813          .
## cat1Colon Cancer              0.0005        0.0017         -0.0576          .
## cat1Coma                      0.0435        0.0960         -0.2575          .
## cat1Lung Cancer               0.0023        0.0096         -0.1524          .
## cat1MOSF w/Malignancy         0.0723        0.0679          0.0173          .
## cat1MOSF w/Sepsis             0.3205        0.1484          0.3688          .
## surv2md1                      0.5685        0.6072         -0.1954     1.0663
## aps1                         60.7390       50.9335          0.4837     1.1609
##                        eCDF Mean eCDF Max
## distance                  0.2181   0.3183
## transhx                   0.0554   0.0554
## age                       0.0285   0.0703
## scoma1                    0.0346   0.0683
## hrt1                      0.0344   0.0782
## bili1                     0.0244   0.1119
## wtkilo1                   0.0535   0.1252
## cat1ARF                   0.0290   0.0290
## cat1CHF                   0.0261   0.0261
## cat1COPD                  0.0858   0.0858
## cat1Cirrhosis             0.0268   0.0268
## cat1Colon Cancer          0.0012   0.0012
## cat1Coma                  0.0525   0.0525
## cat1Lung Cancer           0.0073   0.0073
## cat1MOSF w/Malignancy     0.0045   0.0045
## cat1MOSF w/Sepsis         0.1721   0.1721
## surv2md1                  0.0475   0.0957
## aps1                      0.0797   0.2127
##
## Summary of Balance for Matched Data:
##                        Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                      0.4379        0.4296          0.0488     1.0901
## transhx                       0.1331        0.1259          0.0202          .
## age                          60.9422       61.0129         -0.0045     0.7935
## scoma1                       18.9162       18.8710          0.0016     0.9685
## hrt1                        118.0288      118.3258         -0.0072     0.9402
## bili1                         2.4512        2.3965          0.0103     0.8783
## wtkilo1                      77.0525       76.2229          0.0384     0.7715
## cat1ARF                       0.4419        0.4687         -0.0542          .
```

```
## cat1CHF                      0.0961        0.1074         -0.0384          .
## cat1COPD                     0.0298        0.0211          0.0543          .
## cat1Cirrhosis                0.0252        0.0272         -0.0139          .
## cat1Colon Cancer             0.0005        0.0000          0.0240          .
## cat1Coma                     0.0483        0.0493         -0.0050          .
## cat1Lung Cancer              0.0026        0.0015          0.0215          .
## cat1MOSF w/Malignancy        0.0755        0.0776         -0.0079          .
## cat1MOSF w/Sepsis            0.2801        0.2472          0.0705          .
## surv2md1                     0.5796        0.5815         -0.0093     0.9341
## aps1                        58.5149       57.7081          0.0398     1.0368
##                       eCDF Mean eCDF Max Std. Pair Dist.
## distance                 0.0105   0.0421          0.0489
## transhx                  0.0072   0.0072          0.6049
## age                      0.0269   0.0601          1.1907
## scoma1                   0.0087   0.0252          0.9512
## hrt1                     0.0088   0.0334          1.0680
## bili1                    0.0053   0.0807          0.5481
## wtkilo1                  0.0187   0.0550          1.0392
## cat1ARF                  0.0267   0.0267          0.8924
## cat1CHF                  0.0113   0.0113          0.6044
## cat1COPD                 0.0087   0.0087          0.2205
## cat1Cirrhosis            0.0021   0.0021          0.3331
## cat1Colon Cancer         0.0005   0.0005          0.0240
## cat1Coma                 0.0010   0.0010          0.4434
## cat1Lung Cancer          0.0010   0.0010          0.0860
## cat1MOSF w/Malignancy    0.0021   0.0021          0.5554
## cat1MOSF w/Sepsis        0.0329   0.0329          0.6166
## surv2md1                 0.0128   0.0391          1.0847
## aps1                     0.0094   0.0344          0.8948
##
## Sample Sizes:
##           Control Treated
## All          3551    2184
## Matched      1946    1946
## Unmatched    1605     238
## Discarded       0       0
```

Matched individuals become less. 1946 treated individuals and 1946 control individuals are matched. After matching, there is no imbalance in covariates.

c

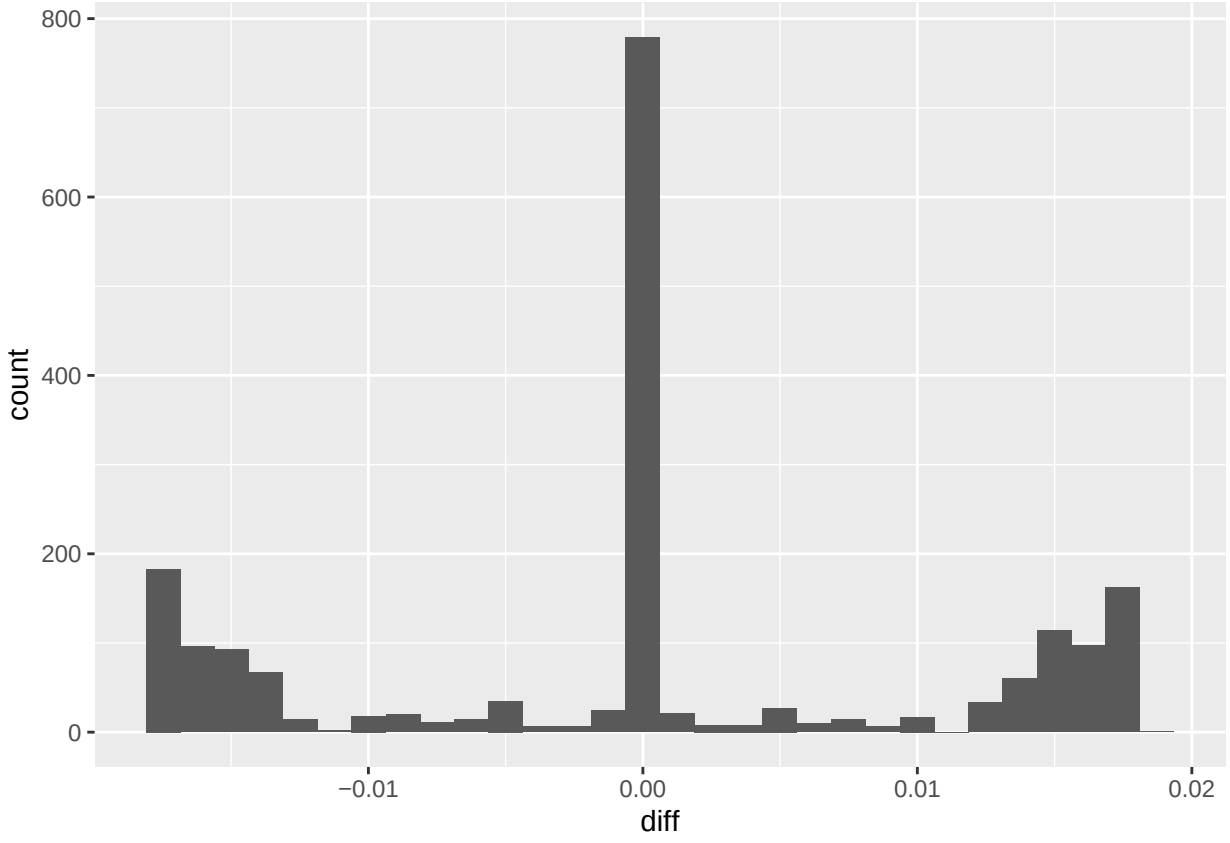
```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##     combine
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
gen_diff = function(data) {
    return(data[1] - data[2])
}
match_rhc = match.data(match_out2)
match_rhc = match_rhc[order(match_rhc$subclass, decreasing = FALSE), ]
ps_diff = summarise(group_by(match_rhc, subclass), diff = gen_diff(distance))
ggplot(data = ps_diff, mapping = aes(x = diff)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As shown on the histogram of propensity score difference within matched pairs, the propensity scores are very similar within matched pairs.

d

```r
match_rhc$death30 = as.numeric(match_rhc$death30) - 1
EY1 = mean(match_rhc[match_rhc$treatment == "RHC", "death30"])
EY0 = mean(match_rhc[match_rhc$treatment == "no RHC", "death30"])
ATE = EY1 - EY0
print(c(EY1 = EY1, EY0 = EY0, ATE = ATE))
```

```
##        EY1        EY0        ATE
## 0.36742035 0.32785200 0.03956835
```

ATE is 0.040 which is larger than the result 0.039 of 2.j.