Quant_II_hwk_06

吴温泉

目录

1 倾向值匹配法：分析共产党员对于收入的因果效用

1.1 使用 CGSS2010 数据，处理变量为 “是否为共产党员”，结果变量为 “个人去年总收入”，共变量为性别、年龄、民族 (分成 8 类)、教育程度（分为 5 类）、身高、体重、说英语能力、说普通话能力、家庭总收入、父亲教育程度（分为 5 类）、父亲是否为共产党员。

```
dat <- read_dta("cgss2010_14.dta")

dat1 <- dat %>% dplyr::select(a10, # ccp member
```

```
                        a8a, # personal income
                        a2, # male
                        a3a, # age
                        a4, # nation
                        a7a, # education
                        a13, # height
                        a14, # weight
                        a50, # English speaking
                        a52, # Mandarin speaking
                        a62, # family income
                        a89b, # father ccp member
                        a89c, # father education
                        )

names(dat1) <- c('ccp_member', 'ind_income', 'male', 'age', 'nation', 'educ',
                 'height', 'weight', 'en_speak', 'mand_speak', 'fam_income',
                 'fat_member', 'fat_educ')

dat1 <- dat1 %>% mutate(ccp_member = if_else(ccp_member==1,1,0),
                        fat_member = if_else(fat_member==1,1,0),
                        male = if_else(male==1, 1,
                                       ifelse(male==2, 0, NA)),
                        age = ifelse(age < 17, NA, age),
                        age = 2010 - age,

                        fam_income = ifelse(fam_income > 9999996, NA, fam_income),
                        ind_income = ifelse(ind_income > 9999996, NA, ind_income),

                        nation = ifelse(nation < 0, NA, nation),
                        height = ifelse(height < 0, NA, height),
                        weight = ifelse(weight < 0, NA, weight),
                        en_speak = ifelse(en_speak < 0, NA, en_speak),
                        mand_speak = ifelse(mand_speak < 0, NA, mand_speak),

                        educ = case_when(1 < educ & educ < 4 ~ 'primary',
                                         educ == 4 ~ 'junior',
                                         4 < educ & educ< 9 ~ 'senior',
                                         8 < educ & educ< 14 ~ 'higher',
```

```
                                    educ == 14 ~ NA_character_,
                                    educ == -3 ~ NA_character_,
                                    educ == 1 ~ 'uneducated'),

                    fat_educ = case_when(1 < fat_educ & fat_educ < 4 ~ 'primary',
                                    fat_educ == 4 ~ 'junior',
                                    4 < fat_educ & fat_educ< 9 ~ 'senior',
                                    8 < fat_educ & fat_educ< 14 ~ 'higher',
                                    fat_educ == 14 ~ NA_character_,
                                    fat_educ == -3 ~ NA_character_,
                                    fat_educ == 1 ~ 'uneducated'),

                    log_ind_income = log(ind_income+1),
                    log_fam_income = log(fam_income+1)
                     )
```

1.2 使用 logit 回归估计倾向值。

```
dat2 <- na.omit(dat1)
dat2$ccp_member <- as.logical(dat2$ccp_member)
attach(dat2)

m1 <- glm(ccp_member ~ male + age  + nation + educ +
  height + weight + en_speak + mand_speak + log_fam_income, family = binomial, data = dat2)

pm1 <- Match(Y = log_ind_income, Tr = ccp_member, X = m1$fitted, estimand = "ATT", M = 1, replace =
```

1.3 使用 1 对 1 最近邻对照组样本可以替换匹配法进行匹配，并估计实验组处理效用 (ATT)。

```
m1 <- glm(ccp_member ~ male + age  + nation + educ +
  height + weight + en_speak + mand_speak + log_fam_income, family = binomial, data = dat2)

pm1 <- Match(Y = log_ind_income, Tr = ccp_member, X = m1$fitted, estimand = "ATT", M = 1, replace =
summary(pm1)
```

```
##
## Estimate...  0.26575
## AI SE......  0.078793
## T-stat.....  3.3727
## p.val......  0.00074429
##
## Original number of observations..............  9312
## Original number of treated obs...............  1229
## Matched number of observations...............  1229
## Matched number of observations  (unweighted).  15792
```

实验组平均处理效用为 0.26575，标准误为 0.078793，在 95% 置信水平下显著。

1.4 使用 1 对 5 最近邻对照组样本可以替换匹配法进行匹配，并估计实验组处理效用。

```
m2 <- glm(ccp_member ~ male + age  + nation + educ +
  height + weight + en_speak + mand_speak + log_fam_income, family = binomial, data = dat2)

pm2 <- Match(Y = log_ind_income, Tr = ccp_member, X = m1$fitted, estimand = "ATT", M = 5, replace =
summary(pm2)
```

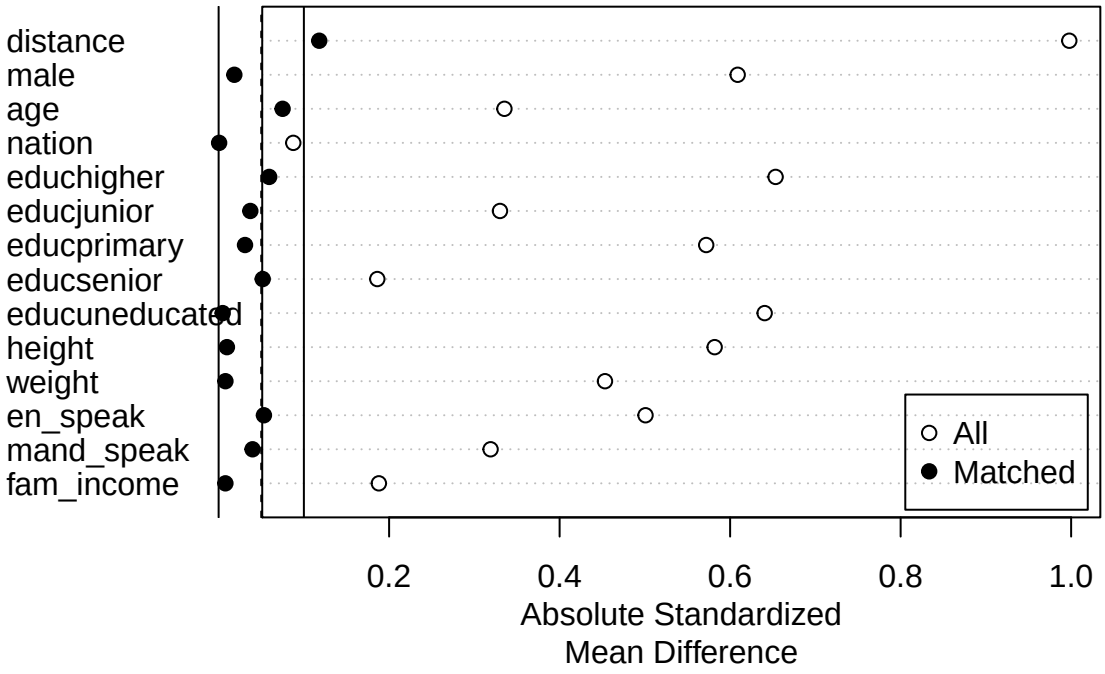
```
##
## Estimate...  0.28565
## AI SE......  0.071821
## T-stat.....  3.9773
## p.val......  0.000069708
##
## Original number of observations..............  9312
## Original number of treated obs...............  1229
## Matched number of observations...............  1229
## Matched number of observations  (unweighted).  18410
```

实验组平均处理效用为 0.28565，标准误为 0.071821，在 95% 的置信水平下显著

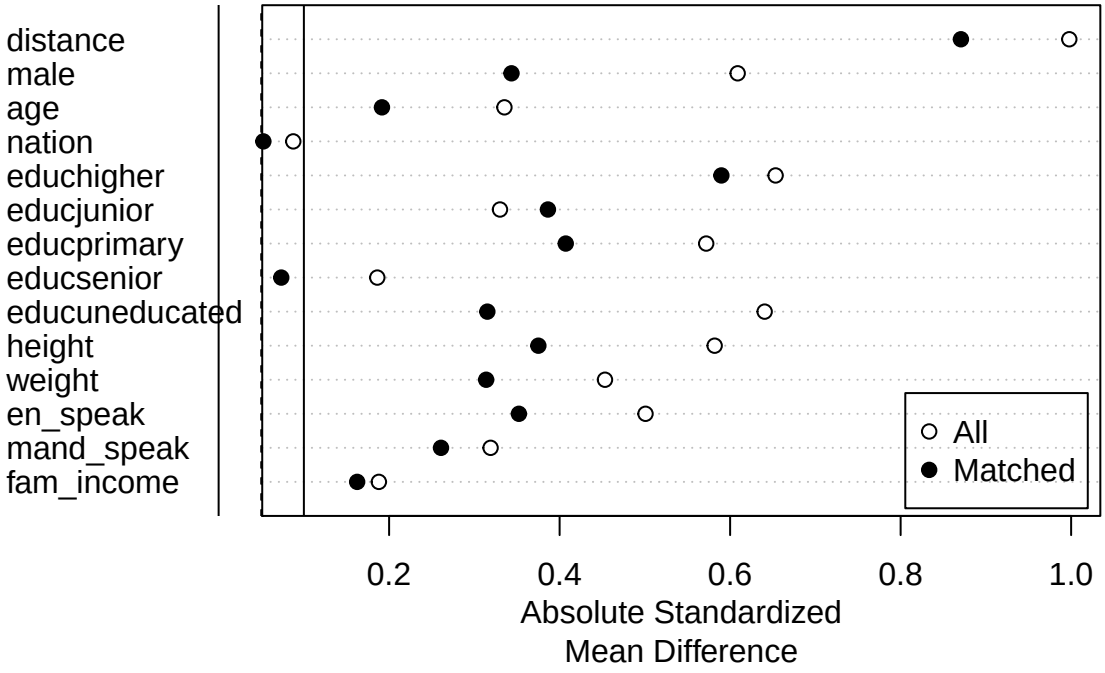
1.5 （平衡和重合检验）比较上面两个方法共变量平衡的情况，依照平衡情况选择较好的匹配模型，重新进行匹配，并（在 R 函数里）加上重合选项。

```
# 平衡性检验

mNearest1v1 <- matchit(ccp_member ~ male + age  + nation + educ +
  height + weight + en_speak + mand_speak + fam_income, data = dat2, method = "nearest", ratio=1)
sNearest1v1 <- summary(mNearest1v1, standardize = TRUE)
plot(sNearest1v1)
```

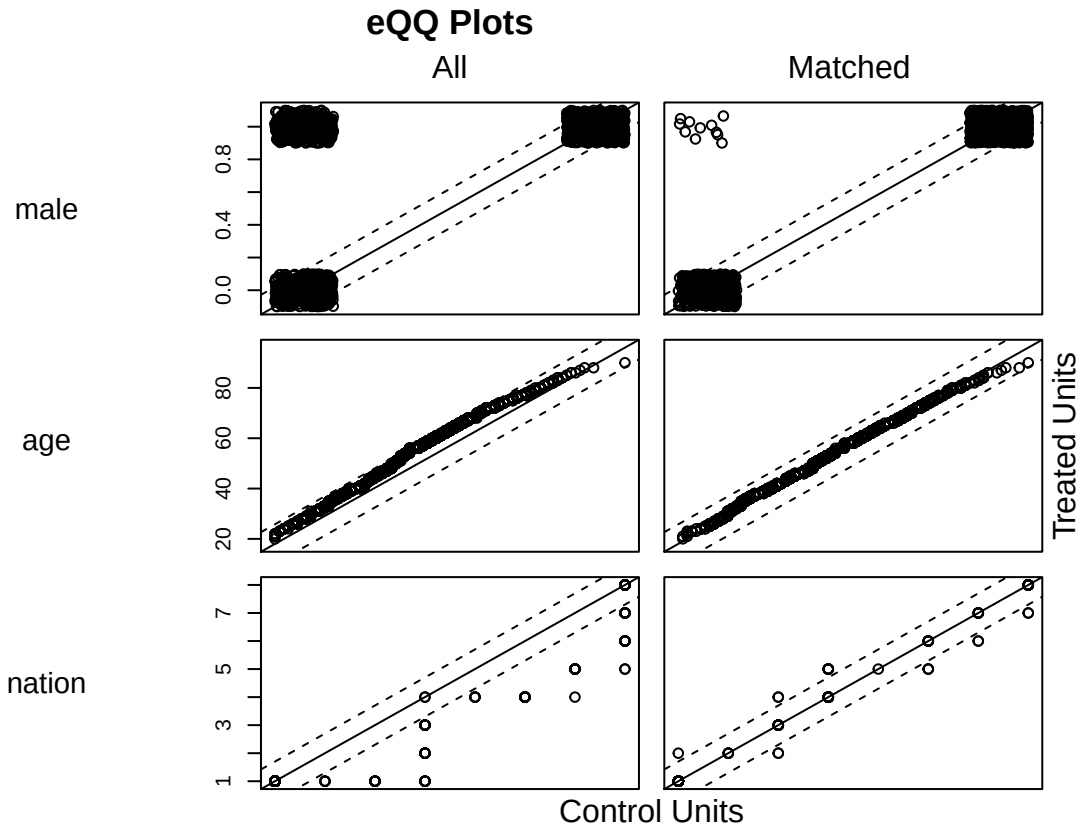


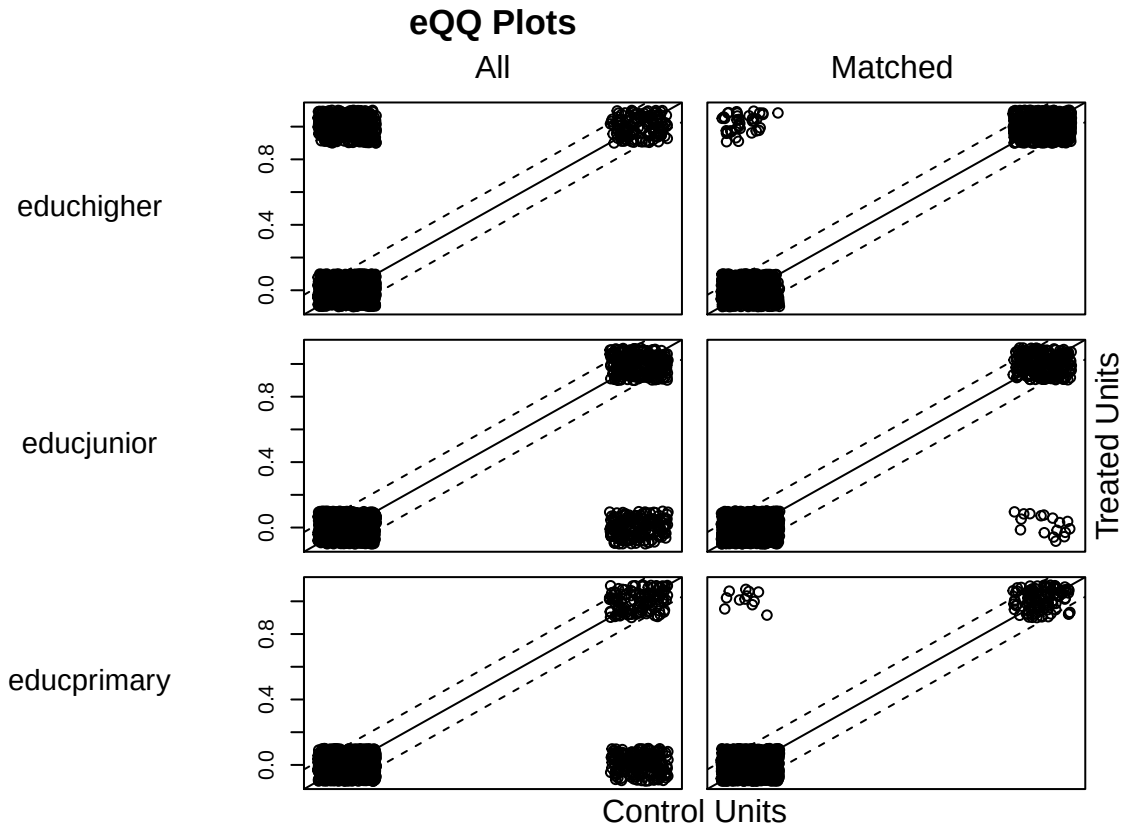

```
mNearest1v5 <- matchit(ccp_member ~ male + age  + nation + educ +
  height + weight + en_speak + mand_speak + fam_income, data = dat2, method = "nearest", ratio=5)
sNearest1v5 <- summary(mNearest1v5, standardize = TRUE)
plot(sNearest1v5)
```

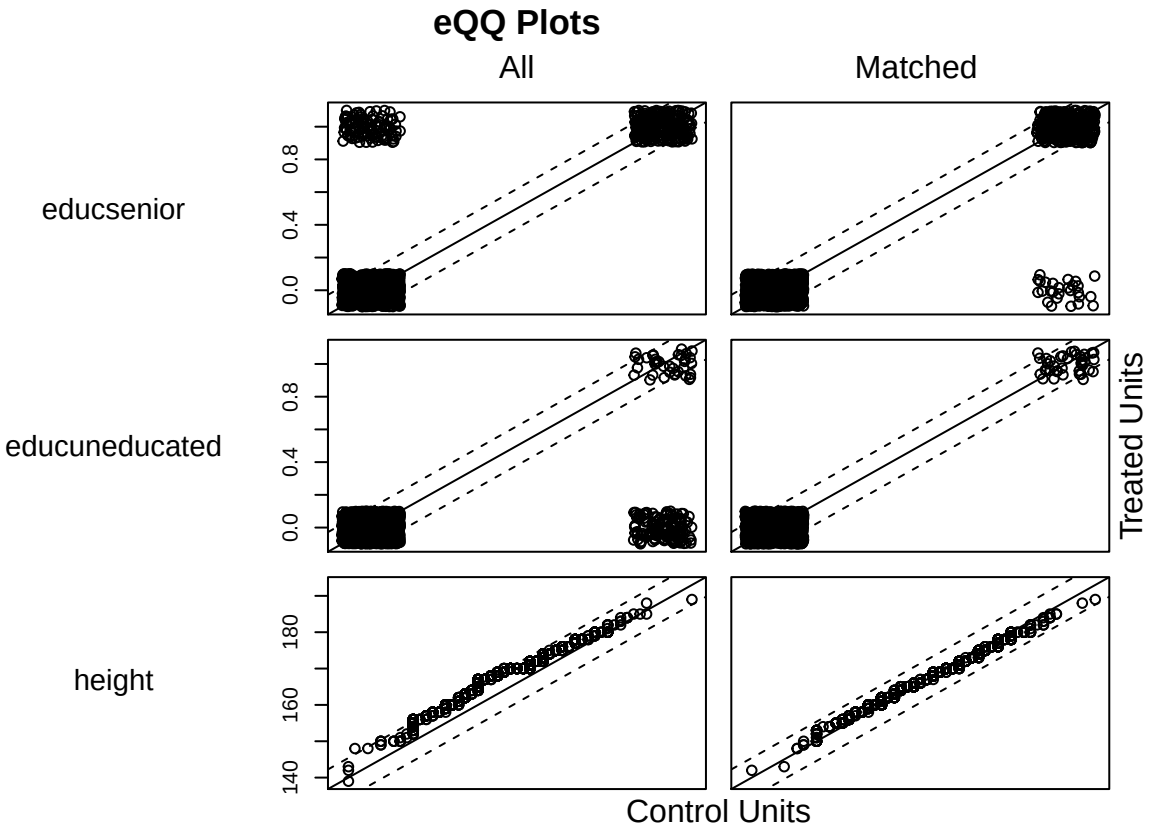


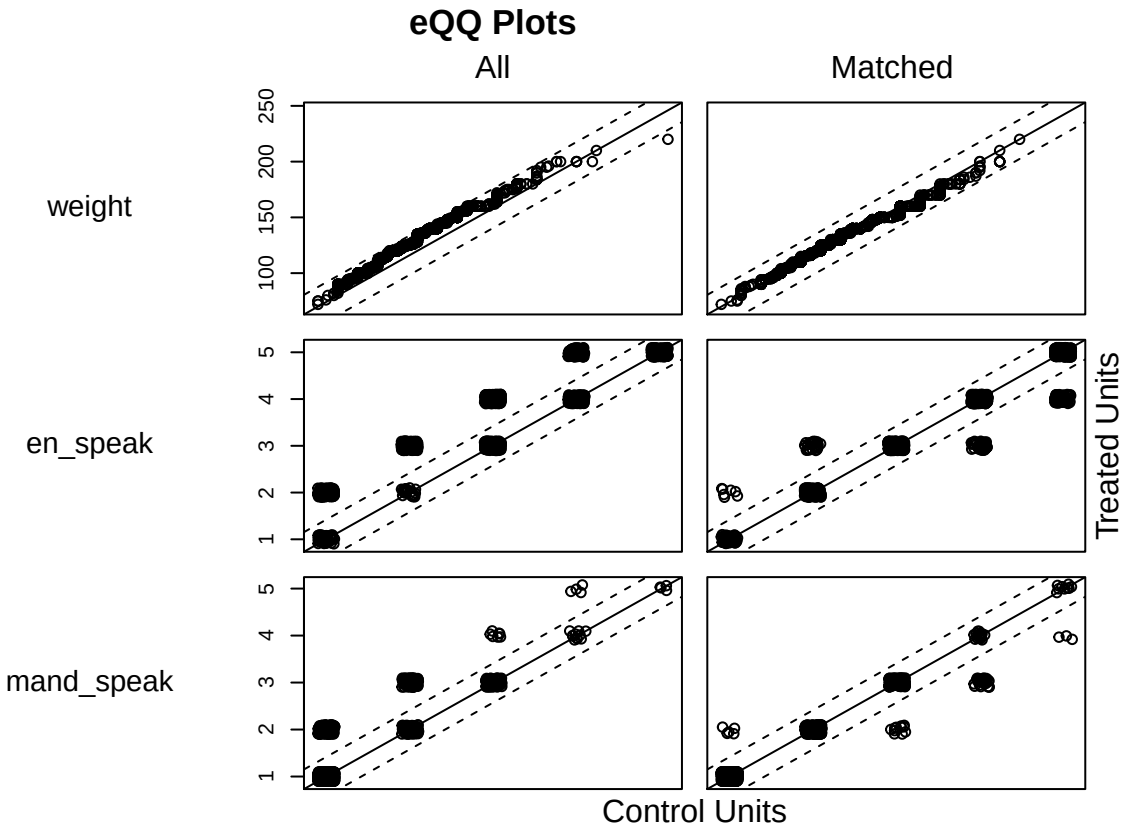
如图所示，1V1 的效果明显更好，更加平衡。

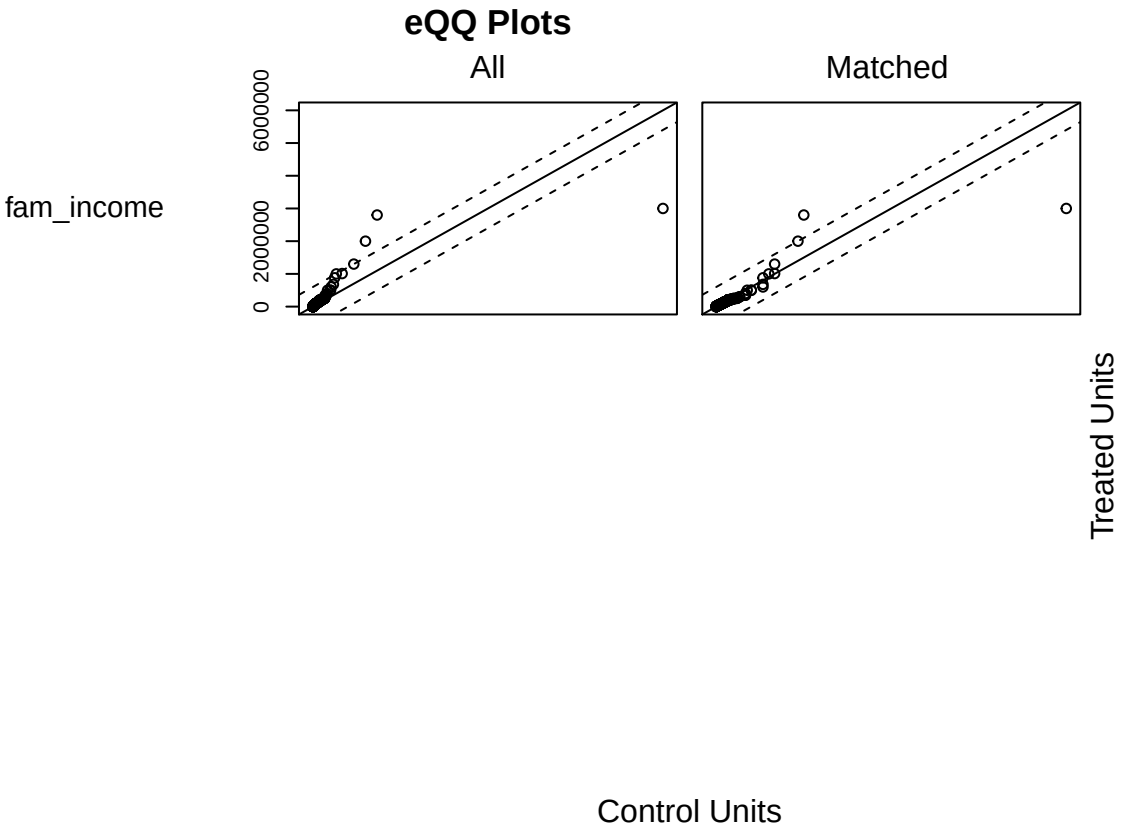
```
# 重合性检验
plot(mNearest1v1, type="QQ")
```


eQQ Plots
All
Matched
male
age
nation
0.0
0.4
0.8
20
40
60
80
1
3
5
7
Treated Units
Control Units

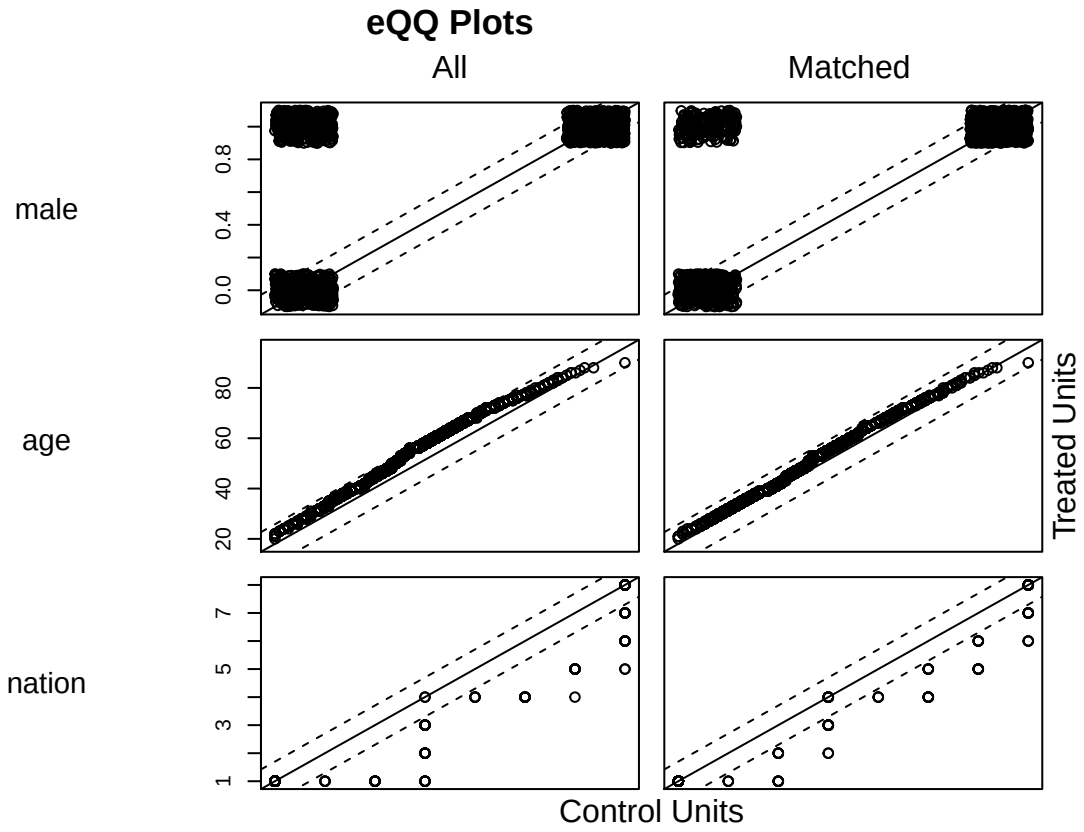

eQQ Plots
All
Matched
educhigher
educjunior
educprimary
0.0
0.4
0.8
Treated Units
Control Units

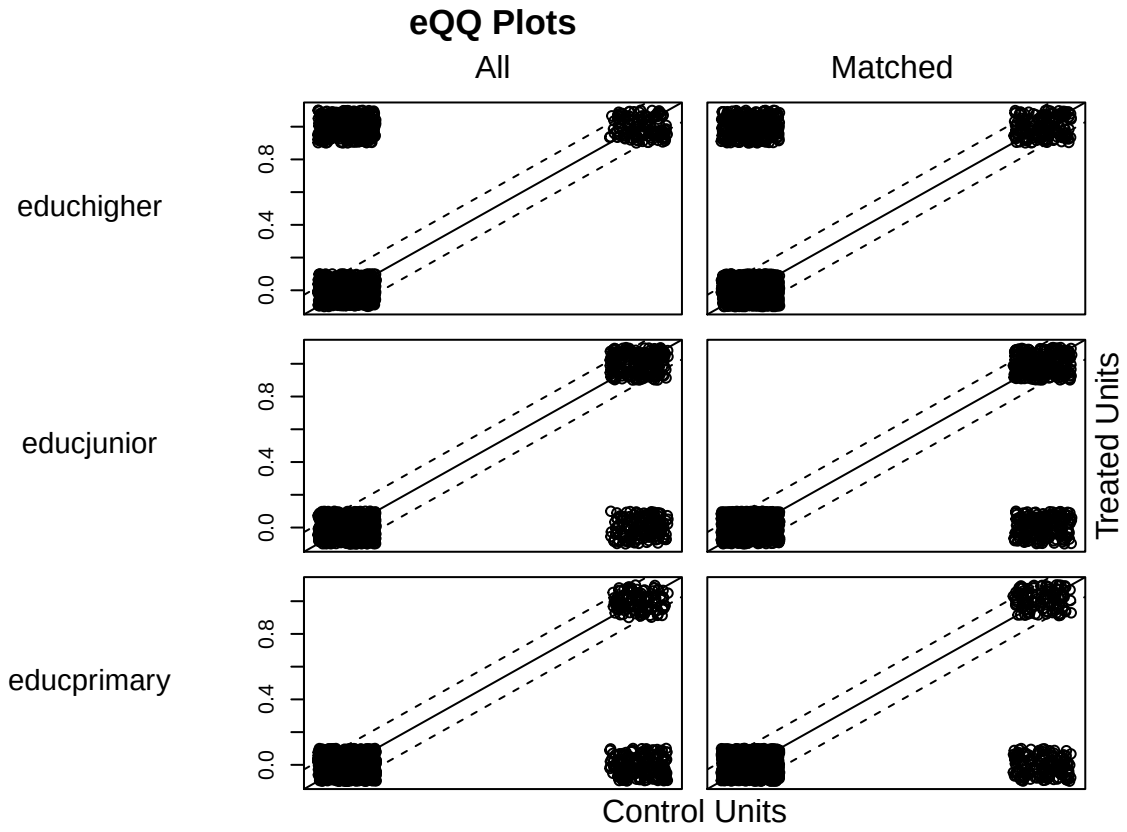

eQQ Plots
All
Matched
educsenior
educuneducated
height
0.0
0.4
0.8
140
160
180
Treated Units
Control Units

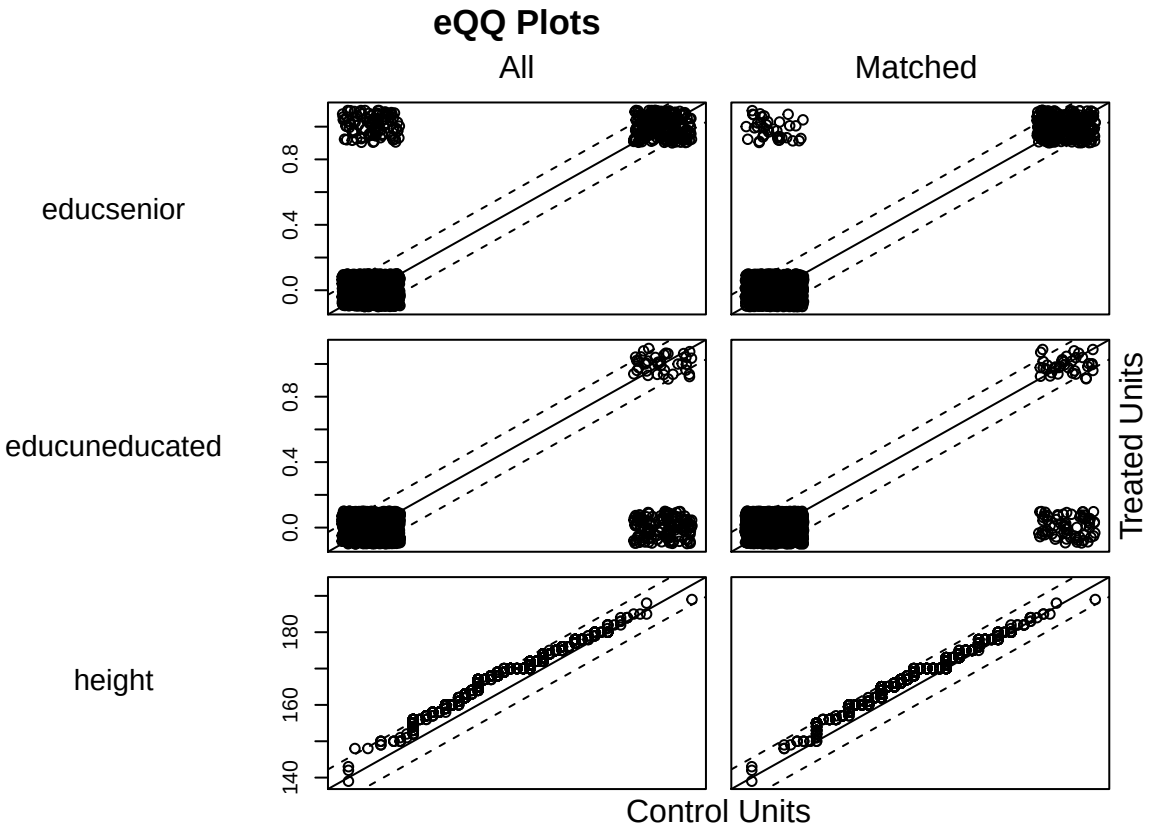

eQQ Plots
All
Matched
weight
en_speak
mand_speak
100
150
200
250
1
2
3
4
5
Treated Units
Control Units

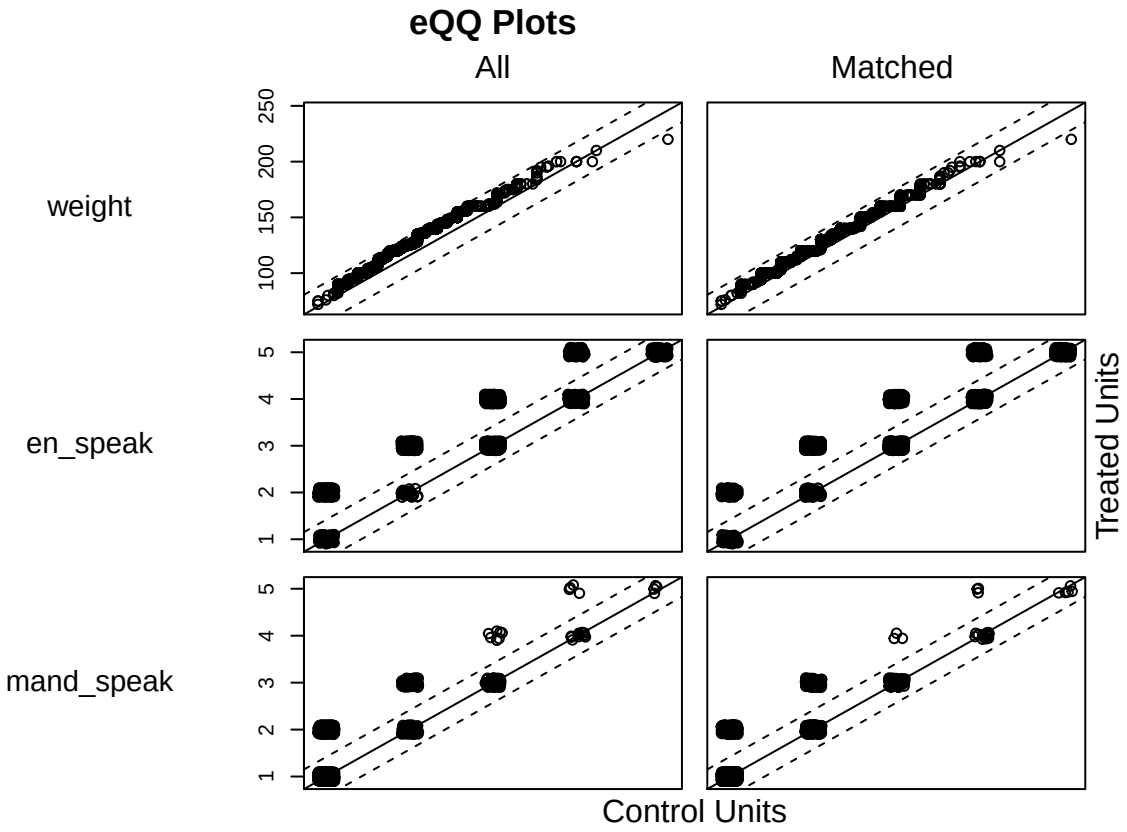


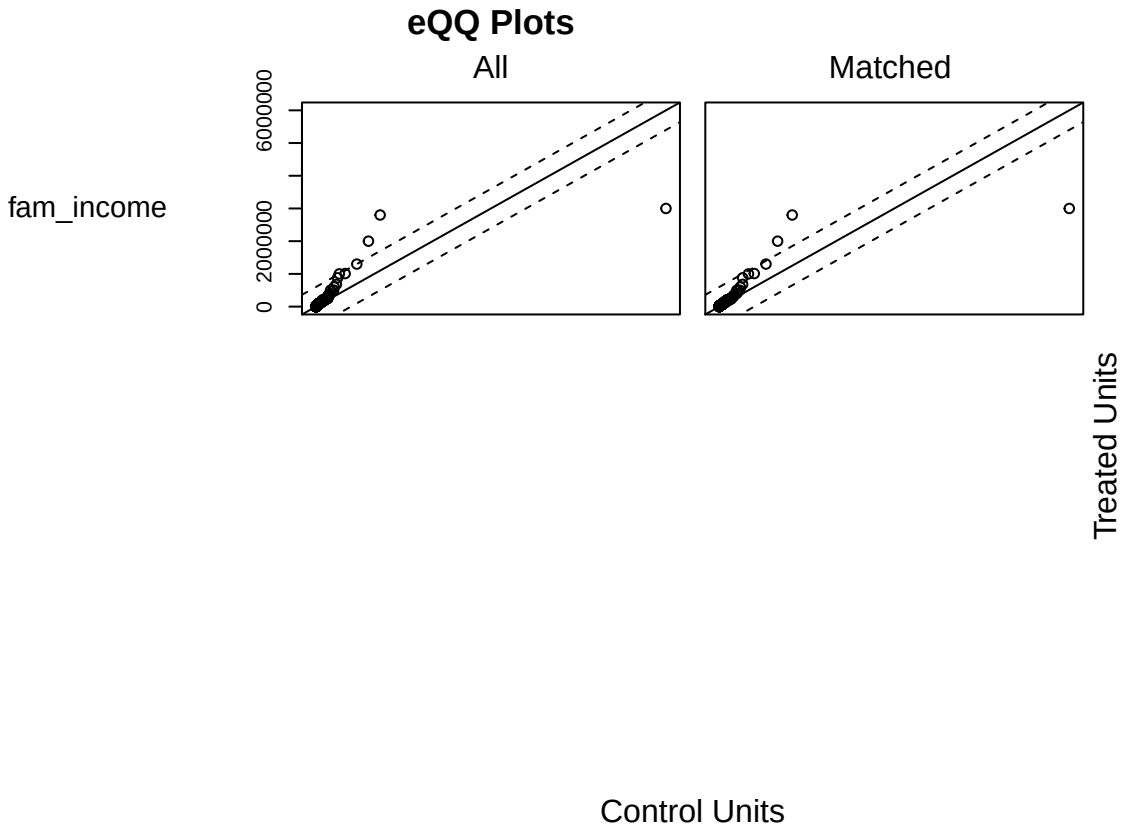

```
plot(mNearest1v5, type="QQ")
```


eQQ Plots
All
Matched
male
age
nation
0.0
0.4
0.8
20
40
60
80
1
3
5
7
Treated Units
Control Units


eQQ Plots
All
Matched
educhigher
educjunior
educprimary
0.0
0.4
0.8
Treated Units
Control Units


eQQ Plots
All
Matched
educsenior
educuneducated
height
0.0
0.4
0.8
140
160
180
Treated Units
Control Units


eQQ Plots
All
Matched
weight
en_speak
mand_speak
100
150
200
250
1
2
3
4
5
Treated Units
Control Units



从 nation，english speak, mandarian speak, log family income 等结果来看，仍是 1V1 的结果更佳。

```
# 加上重合选项
csMatch <- Match(Y = log_ind_income, Tr = ccp_member, X = m1$fitted, estimand = "ATT", CommonSuppo
summary(csMatch)
```

```
##
## Estimate...  0.26699
## AI SE......  0.078938
## T-stat.....  3.3823
## p.val......  0.00071887
##
## Original number of observations..............  8688
## Original number of treated obs...............  1224
## Matched number of observations...............  1224
## Matched number of observations  (unweighted).  15846
```

结果好像也没有更好。

1.6 依照以上分析结果，选择最适合的匹配结果，进行敏感性分析，并说明处理效用是否通过敏感性检验。

```
psens(x = pm1, Gamma = 2, GammaInc = 0.1)
```

```
##
##  Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-Value
##
## Unconfounded estimate ....  0
##
##  Gamma Lower bound Upper bound
##    1.0           0      0.0000
##    1.1           0      0.0000
##    1.2           0      0.0032
##    1.3           0      0.9351
##    1.4           0      1.0000
##    1.5           0      1.0000
##    1.6           0      1.0000
##    1.7           0      1.0000
##    1.8           0      1.0000
##    1.9           0      1.0000
##    2.0           0      1.0000
##
##  Note: Gamma is Odds of Differential Assignment To
##  Treatment Due to Unobserved Factors
##
```

```
hlsens(x = pm1, Gamma = 2, GammaInc = 0.1)
```

```
##
##  Rosenbaum Sensitivity Test for Hodges-Lehmann Point Estimate
##
## Unconfounded estimate ....  0.195
##
##  Gamma Lower bound Upper bound
##    1.0   0.1949500     0.19495
##    1.1   0.0949540     0.29495
##    1.2  -0.0050456     0.39495
```

```
##    1.3  -0.1050500     0.39495
##    1.4  -0.1050500     0.49495
##    1.5  -0.2050500     0.59495
##    1.6  -0.2050500     0.59495
##    1.7  -0.3050500     0.69495
##    1.8  -0.3050500     0.69495
##    1.9  -0.4050500     0.79495
##    2.0  -0.4050500     0.79495
##
##  Note: Gamma is Odds of Differential Assignment To
##  Treatment Due to Unobserved Factors
##
```

Wilcoxon 符号秩检验 P 值在 1.3 时不显著，低于 2，未通过检验。

Hodges-Lehmann 检验法点估计在 1.2 时符号不同，未通过检验。

1.7 请以文字说明共产党员对于收入的因果效用为何，并说明整个分析过程中可能违反因果推论假设和要求的部分？

共产党员对于个人收入的处理效用为 0.26575。可能违反违反因果推论假设和要求的部分有：

1. 未通过敏感性检验，可能存在遗漏变量偏差
2. 存在缺失值
3. 未进行 bootstrap
4. 家庭收入、教育程度、民族等变量在某些区间重合度不高