

# Lalonde: Matching

CS112

Tomer Eldor 21/03/2017

## Comparison of Results

	Treatment Effect	Confidence Interval: Lower end   Upper end		After Matching Lowest P Values	Notes
1.Experiment	<b>1794.342</b>	550.5745	3038.110		Significantly positive
2.Observational (cps) ATE	-8497.516	-9893.156	7101.877		Significantly negative; but a wrong comparison to make
3. Propensity Scores Matching	1732.9	-193.7	3659.5	$p < 2.22e-16$ ; for 8 variables (vs. 11 variables before matching)	Not statistically significant, didn't improve balance
4.1 Multivariate Matching – with nonlinear terms	1479	-138	3096	$p=0.07$ for 2 variables (9,10)	Not statistically significant
4.2 Multivariate Matching – with only linear terms	1972.5 (with caliper=0.24; whereas no caliper yielded 2071)	169	3450	$p=0.013752$ for 1 variable (propensities)	Positive (statistically significant, but barely).
5.1. GenMatch Propensity Scores	1937.9	-9	3885	$p < 2.22e-16$ for 1-2 variables	Here we used only propensity scores.
5.2. GenMatch Multivariate Matching	1850.969 *used with caliper=0.45	394.6895	3307.249	0.05, for only 1 variable (re74). Best covariance balance achieved with caliper=0.45	Amongst the best p values (but surprisingly, 4.1 was higher), but the closest estimand to the experiment.

## Summary

These experiments show the effectiveness and the pitfalls for various methods to get treatment effect estimates, including several matching approaches.

We see that Multivariate Matching and Genetic Matching yield the best balance and the closest results to the experimental results.

The observational Average Treatment Effect proves to be the worst estimate. That makes sense, since without matching, in a sample size so large (11,000+ observations) there are for sure underlying inherent differences and outliers which make the comparison unbalanced.

While multivariate matching with nonlinear terms achieved better balance, multivariate matching with only linear terms yielded results closer to the experimental values.

### Propensity Score Matching

Before Matching Minimum p.value: < 2.22e-16

Variable Name(s): age I(age^2) education I(education^2) black married  
nodegree re74 I(re74^2) re75 I(re75^2)

After Matching Minimum p.value: < 2.22e-16

Variable Name(s): age I(age^2) education I(education^2) re74 I(re74^2) re75  
I(re75^2)

We didn't increase our smallest p-value, but we did reduce the number of variables with small p-values (we improved the p-values for black, married, and nodegree). This Matching increased many of p-values, and also the var ratio (Tr/Co) to be more balanced and closer to 1.

This is because propensity scores matching only considers one-dimensional space of the "propensity score" to match on; so, we lose the valuable information of WHICH covariates affected the "chance to be in the treatment group"; and we don't match units who are similar in many parameters, but only in the sense of "how likely they are to be in the treatment group".

### Genetic Matching with Propensity Scores

We see practically no improvement in the matching balance. The matching didn't achieve better balance than before: not in the lower p-value, and not in the number of variables with that low p-value.

This is because the GenMatch() is meant to get the weights for all the covariates, but if we end up only matching on propensity scores, then it does nothing in practice.

The Genetic Matching Function actually only considers ONE COVARIATE, the propensity scores; therefore, it doesn't do its job of optimizing different weights for different parameters; it assigns everything onto that one covariate, and just checks what's the optimal covariate for it.

We can see that demonstrated in the resulting weight matrix from the `genmatch`, which contains only one weight!

### Genetic Multivariate Matching

As expected, this yielded the best results.

Exploring the parameters, the default parameters for caliper yielded the best results; best caliper used was 0.45.

## 6: Experimenting with the `genmatch` and `match` functions

### Caliper

I've experimented with the caliper in `Match` and `GenMatch` functions, as talked about more in length in question 4. Starting from `Caliper=0.0` (exact matching), I've slowly, iteratively, increased the caliper to find a balanced point where not too many observations are dropped. What happens is that the caliper (or "maximum difference distance between matched units" allowed), if set to 0 or some very small amount, will find only exactly or very closely matched units, and drop the rest. Therefore, it might change the underlying structure of the treated group taken into consideration - for example, it may only find good matches to Black, 30-year-old people with No Degree. That would decrease our external validity and generalizability, since our results would only apply to that subpopulation. In the cases I checked here, it seemed that usually, when I set the caliper to 0.0, the resulting ATT was further away from the experimental results (ATE). As the caliper increased approaching the point of dropping only ~10% of units or less, I saw that the Matched results approach the experimental results. This happens since if we limit the compared subpopulations too much, we miss information. We don't want to eliminate much of our precious "treatment" unit's information, and want to consider a wider set of treated samples so we'd have more external validity and generalizability. However, for genetic matching, increasing the caliper incrementally until even over 1.3, the performance of the balance and matching was worse than without specifying caliper. When Genetic Matching doesn't use caliper, it uses distance. Tolerance, which performed better.

### Replace

`Replace` sets whether matching should be done with replacement or not. When I set `replace=False`, that reduced the quality of the results, resulting in poorer resulting matching - most p-values were not increased by as much. This happens since if we remove observations from the pool every time we match them, we decrease the pool size from which we can choose the best match for the next units. Therefore, that same observation which was dropped might have been the best match for next 3 units, but we will have to suffice with increasingly worse matches each time. So usually `Replace` decreases the quality of matching, and this was visible in this example by not increasing as much many of the P-Values.

## Pop Size, Generations

The Population Size and Generations each determine how deep / wide should our genetic search be. Therefore, increasing these improves our result, but takes considerably more time. As I increased the Population size and generations of GenMatch, it took much longer, but overall normally improved the results - seeing with better improved p-values, and sometimes even a smaller AI Standard Error (although I'm not sure it relates specifically to this).

## Estimand

Estimand: we are interested in estimating the ATT - the sample average treatment effect for the treated. Otherwise the estimand could be: "ATE" - the sample average treatment effect, or "ATC" - the sample average treatment effect for the controls.

Using ATE *completely* changed the results, to a NEGATIVE, but insignificant effect. Estimate... -509.14 AI SE..... 495.11 The confidence interval crosses 0 to the positive side. But this is not the measure of our interest, since we want the treatment effect FOR THE TREATED; therefore, we want to take the treated units and match controls onto them, measuring the difference between these pairs. But ATE considers the entire control group, instead of focusing on the treated, therefore we have differences which aren't most representative of the actual effect of our treatment.

We also didn't improve our balance by much this way. We remained with 3 variables with the unimproved lowest p-value result.

Before Matching Minimum p.value: < 2.22e-16 Variable Name(s): propensities age education black married nodegree re75 re74 Number(s): 1 2 3 4 6 7 8 9

After Matching Minimum p.value: < 2.22e-16 Variable Name(s): propensities age education re75 re74 Number(s): 1 2 3 8 9

## M

M=1 is one-to-one matching, which is also the default. Increasing this will make matching into many-to-one matching. In my experiments, increasing the M DECREASED the quality of covariance balance after Matching. Increasing the M up to 5, decreased the improvement in p-value after matching up to the point of having no improvement in the lowest p-value. Moreover, the resulting estimate was *significantly far away* from the experimental results and even from any other method: Estimate... 1014.2 AI SE..... 620.31 .

## DETAILED Process & Discussion

*# Load the Libraries*

```
library(foreign)
```

```
library(Matching)
```

```
## Loading required package: MASS
```

```
## ##
```

```
## ## Matching (Version 4.9-2, Build Date: 2015-12-25)
```

```
## ## See http://sekhon.berkeley.edu/matching for additional documentation.
```

```
## ## Please cite software as:
```

```
## ## Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score Matching
```

```
## ## Software with Automated Balance Optimization: The Matching package fo
```

```
r R.''
```

```
## ## Journal of Statistical Software, 42(7): 1-52.
```

```
## ##
```

```
library(rgenoud)
```

```
## ## rgenoud (Version 5.7-12.4, Build Date: 2015-07-19)
```

```
## ## See http://sekhon.berkeley.edu/rgenoud for additional documentation.
```

```
## ## Please cite software as:
```

```
## ## Walter Mebane, Jr. and Jasjeet S. Sekhon. 2011.
```

```
## ## ``Genetic Optimization Using Derivatives: The rgenoud package for R.'
```

```
,
```

```
## ## Journal of Statistical Software, 42(11): 1-26.
```

```
## ##
```

*# Load the data*

```
dw <- read.dta("nsw_dw.dta")
```

## 1: Simple Average Treatment Effect in Randomized Experiment

Calculating the simple average treatment effect: the difference of means

*#Treatment Effects: a simple difference in means*

```
TEdw <- mean(dw[dw$treat == 1,]$re78) - mean(dw[dw$treat == 0,]$re78)
```

```
#mean(dw[dw$treat == 1,]$re78) ; mean(dw[dw$treat == 0,]$re78)
```

```
TEdw
```

```
## [1] 1794.342
```

The overall ATE average treatment affect is 1794.342.

**confidence intervals for the results:**

*#confidence interval by running a univariate linear regression (with just the treat variable)*

```
lm1 <- lm(re78 ~ treat, data=dw)
```

```
confint(lm1, level=0.95) ; #summary(lm1)
```

```
##              2.5 %   97.5 %
## (Intercept) 3752.8550 5356.747
## treat       550.5745 3038.110
```

Confidence interval for simple ATE of 95% is: 550.5749 to 3038.111.

for the overall ATE average treatment affect is 1794.342, and a 95% confidence interval shows we are 95% sure that the true treatment effect will be between 550.5749 and 3038.111, which is an (undesirably) a very wide range; the low end would show little effect, the upper end is actually close to the entire control groups' mean (\$4554), meaning that the true effect could have been close to increasing their earnings by 1.5 (while the lower end is only marginally significant).

## 2: Average Treatment Effect With Observational Sample as Control

```
#load cps_controls.dta as control group
cps_controls <- read.dta("cps_controls.dta")
treated <- dw[dw$treat == 1,]

#Treatment Effects: a simple difference in means

TE.cps <- mean(treated$re78) - mean(cps_controls$re78)
#mean(treated$re78)
#mean(cps_controls$re78)
TE.cps

## [1] -8497.516

#COMBINE DW treated group to cps group into one dataset
# It's helpful to use DW here since it has re74 data, like cps (unlike NSW)
 #(cpscombined <- cps_controls)

#checking they have the same columns and underlying data structure
#names(cps_controls) ; names(dw)
#head(cps_controls) ; head(dw)

cpscombined <- rbind(treated,cps_controls)

# convert {1/0}s to TRUE and FALSE, necessary for later used functions:
cpscombined$treat <- cpscombined$treat == 1
```

The overall ATE average treatment affect is -8497.516, which is worryingly negative.

### Confidence intervals for the results:

```
#confidence interval by running a univariate linear regression (with just the
treat variable)
lm2.cps <- lm(re78 ~ treat, data=cpscombined)
confint(lm2.cps, level=0.95)
```

```
##                2.5 %    97.5 %
## (Intercept) 14697.411 14995.908
## treatTRUE   -9893.156 -7101.877
```

```
#summary(Lm2.cps)
```

The simple average treatment effect between the treated from the experiment and the cps\_controls seems extremely negative : -8497.516, and with 95% confidence interval would be found within -9893.156 to -7101.877 (which is relatively close effect).

If this would be an actually representative measure of comparing well-balanced groups, it would hint that our program might not only be "not benefitting", but actually HARMING, compared to the alternatives that our control group had. However, we must not jump to conclusions quickly from this superficial measure. Looking at the cps control 78 income, they are overall much higher. It seems that there might be inherent differences between the groups and that the cps group is a group with higher earnings anyway; therefore, they are not suitable for comparison; they aren't well balanced or matching our treatment groups, thus are not suitable to serve as their counterfactuals. Additionally, there are other measures to be made, for example - median or quantile regression, that might show us different measures, for example, if there are a few outliers which drag the average severely

### 3: PROPOENSITY SCORE MATCHING

Using Propensity Score Matching to produce an estimated treatment effect and confidence interval.

#### 3.1 Estimating Balance of key covariates on the dataset before

```
# running balance tests (t-tests) for key covariates on the dw data before m  
atching (FYI: the results show balance)
```

```
attach(cpscombined)
```

```
MatchBalance(treat~ age + black + hispanic + nodegree + re75 + re78, data=cps  
combined, nboots=500)
```

```
##
## ***** (V1) age *****
## before matching:
## mean treatment..... 25.816
## mean control..... 33.225
## std mean diff..... -103.55
##
## mean raw eQQ diff..... 7.4865
## med  raw eQQ diff..... 6
## max  raw eQQ diff..... 16
##
## mean eCDF diff..... 0.18628
## med  eCDF diff..... 0.19692
## max  eCDF diff..... 0.34274
##
## var ratio (Tr/Co)..... 0.41964
```

```

## T-test p-value..... < 2.22e-16
## KS Bootstrap p-value.. < 2.22e-16
## KS Naive p-value..... 0
## KS Statistic..... 0.34274
##
##
## ***** (V2) black *****
## before matching:
## mean treatment..... 0.84324
## mean control..... 0.073537
## std mean diff..... 211.13
##
## mean raw eQQ diff..... 0.76757
## med  raw eQQ diff..... 1
## max  raw eQQ diff..... 1
##
## mean eCDF diff..... 0.38485
## med  eCDF diff..... 0.38485
## max  eCDF diff..... 0.76971
##
## var ratio (Tr/Co)..... 1.9506
## T-test p-value..... < 2.22e-16
##
##
## ***** (V3) hispanic *****
## before matching:
## mean treatment..... 0.059459
## mean control..... 0.072036
## std mean diff..... -5.3038
##
## mean raw eQQ diff..... 0.016216
## med  raw eQQ diff..... 0
## max  raw eQQ diff..... 1
##
## mean eCDF diff..... 0.0062883
## med  eCDF diff..... 0.0062883
## max  eCDF diff..... 0.012577
##
## var ratio (Tr/Co)..... 0.84109
## T-test p-value..... 0.47458
##
##
## ***** (V4) nodegree *****
## before matching:
## mean treatment..... 0.70811
## mean control..... 0.29584
## std mean diff..... 90.437
##
## mean raw eQQ diff..... 0.41081
## med  raw eQQ diff..... 0

```



```
## max   raw eQQ diff..... 1
##
## mean  eCDF diff..... 0.20614
## med   eCDF diff..... 0.20614
## max   eCDF diff..... 0.41227
##
## var ratio (Tr/Co)..... 0.99753
## T-test p-value..... < 2.22e-16
##
##
## ***** (V5) re75 *****
## before matching:
## mean treatment..... 1532.1
## mean control..... 13651
## std mean diff..... -376.45
##
## mean raw eQQ diff..... 12112
## med   raw eQQ diff..... 13837
## max   raw eQQ diff..... 22438
##
## mean eCDF diff..... 0.4751
## med   eCDF diff..... 0.51248
## max   eCDF diff..... 0.6509
##
## var ratio (Tr/Co)..... 0.12059
## T-test p-value..... < 2.22e-16
## KS Bootstrap p-value.. < 2.22e-16
## KS Naive p-value..... 0
## KS Statistic..... 0.6509
##
##
## ***** (V6) re78 *****
## before matching:
## mean treatment..... 6349.1
## mean control..... 14847
## std mean diff..... -108.01
##
## mean raw eQQ diff..... 9094.9
## med   raw eQQ diff..... 10672
## max   raw eQQ diff..... 34743
##
## mean eCDF diff..... 0.34623
## med   eCDF diff..... 0.36368
## max   eCDF diff..... 0.4714
##
## var ratio (Tr/Co)..... 0.66503
## T-test p-value..... < 2.22e-16
## KS Bootstrap p-value.. < 2.22e-16
## KS Naive p-value..... 0
## KS Statistic..... 0.4714
```

```
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): age black nodegree re75 re78 Number(s): 1 2 4 5 6
```

Shows very low p-values, many of them even around 2.22e-16. P-values are low and most var ratios between (Treatment/Control) are very low. This suggests we should indeed use Matching.

Before Matching Minimum p.value: < 2.22e-16 Variable Names: age black nodegree re75 re78

### 3.2 propensity score matching

*# Carrying out propensity score matching*

*# Estimate the propensity model: Fitting Generalized Linear Models to later find weights*

```
glm.cps <- glm(formula = treat ~ age + I(age^2) + education + I(education^2)
+
                black + hispanic + married + nodegree + re74 + I(re74^2)
                + re75 + I(re75^2) , family=binomial, data=cpscombined)
propensities <- glm.cps$fitted.values
#summary(glm.cps)
```

*#save data objects for matching according to glm cps*

```
X <- propensities
Y <- cpscombined$re78
Tr <- cpscombined$treat
```

*# one-to-one matching with replacement (the "M=1" option).*

*# Estimating the treatment effect on the treated (the "estimand" option defaults to ATT).*

*#*

```
rr0 <- Match(Y=Y, Tr=Tr, X=X, M=1, estimand = "ATT", BiasAdj=FALSE, replace=
TRUE, sample = TRUE) #, caliper = 0.24)
```

*#- when using caliper, it gave worse balance and worse results.*

```
summary(rr0)
```

```
##
## Estimate... 1732.9
## AI SE..... 982.99
## T-stat..... 1.7629
## p.val..... 0.077915
##
## Original number of observations..... 16177
## Original number of treated obs..... 185
## Matched number of observations..... 185
## Matched number of observations (unweighted). 3687
```

*#Interpretation: AI SE is matching corrected standard error due to Abadie and Imbens (hence the name AI).*

Finding Confidence Intervals for the estimated treatment effect

```
#estimated treatment effect
est1 = rr0$est ; est1

##           [,1]
## [1,] 1732.936

#confidence intervals
error = 1.96*(982.99) ; error

## [1] 1926.66

confint1.low <- est1 - error ; confint1.up <- est1 + error ;
confint1.low ; confint1.up

##           [,1]
## [1,] -193.7244

##           [,1]
## [1,] 3659.596
```

### Summary of results for (3)

*Without using caliper:*

The resulting treatment effect was: 1732.9 Confidence interval of: [-193.7244 , 3659.596]  
Which crosses zero, therefore the treatment effect is not statistically significant. In other words, this doesn't reject the null hypothesis that our treatment has no significant effect.  
#####With using the best caliper found: The resulting treatment effect was: 1646.766  
With the confidence interval of: [-279.8944 , 3573.426]

### Trying to find optimal caliper.

**Eventually, Not using Caliper proved to be better.** Although the Caliper parameter might be tricky, I tried to explore the Caliper parameter. To optimize the tradeoff between achieving only exact matching (caliper = 0) to matching unsimilar units, I slowly increased the caliper until finding a small enough caliper which still doesn't drop too many units (since we have only 185 treated units initially), while still achieving small SE. When I increased the caliper to 0.24, the number of excluded units reduced from 7 to 3, while the estimands of the results (the Estimate itself, the corrected Standard Error here called AI SE, and the T-stat and P-value) remained low and similar to before.

It's useful to note that changing the caliper size actually changed the resulting estimations, so there is room for subjectivity or researcher-induced variance here.

So this matching resulted in that, for the caliper of 0.24, we excluded only 3 treated units that we didn't find a good counterfactual for. In this matching, the estimated ATT was a

positive **\$1646.8**, with SE of 957.17; For the original number of treated obs, 185 we matched a counterfactual for 182 units (out of a pool of 16177), which were weighted combinations of 3684 units. Matched number of observations..... 167 Matched number of observations (unweighted). 3667

Caliper (SDs)..... 0.1 Number of obs dropped by 'exact' or 'caliper' 18

**However, without caliper we achieved better results.**

Estimate... 1732.9 AI SE..... 982.99 T-stat..... 1.7629 p.val..... 0.077915

Original number of observations..... 16177 Original number of treated obs..... 185

Matched number of observations..... 185 Matched number of observations (unweighted). 3687

### 3.3 Checking the covariate balance

```
# Checking the covariate balance
# 'nboots' is automatically set to small values in the interest of speed, and
# should be increased to at least 500 each for publication quality p-values.
mb <- MatchBalance(treat ~ age + I(age^2) + education + I(education^2) + bla
ck + hispanic + married + nodegree + re74 + I(re74^2) + re75 + I(re75^2),
data=cpscombined, match.out=rr0, nboots=500)

##
## ***** (V1) age *****
##               Before Matching      After Matching
## mean treatment.....      25.816      25.816
## mean control.....      33.225      26.452
## std mean diff.....     -103.55     -8.8848
##
## mean raw eQQ diff.....      7.4865      3.4264
## med  raw eQQ diff.....        6        2
## max  raw eQQ diff.....      16      11
##
## mean eCDF diff.....      0.18628      0.085659
## med  eCDF diff.....      0.19692      0.088554
## max  eCDF diff.....      0.34274      0.22647
##
## var ratio (Tr/Co).....      0.41964      0.86826
## T-test p-value..... < 2.22e-16      0.35777
## KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16
## KS Naive p-value..... < 2.22e-16 < 2.22e-16
## KS Statistic.....      0.34274      0.22647
##
##
## ***** (V2) I(age^2) *****
##               Before Matching      After Matching
## mean treatment.....      717.39      717.39
## mean control.....      1225.9      758.35
## std mean diff.....     -117.92     -9.4962
```

```

##
## mean raw eQQ diff.....      513.91      209.46
## med   raw eQQ diff.....      336      112
## max   raw eQQ diff.....      1312      1045
##
## mean eCDF diff.....      0.18628      0.085659
## med   eCDF diff.....      0.19692      0.088554
## max   eCDF diff.....      0.34274      0.22647
##
## var ratio (Tr/Co).....      0.302      0.98537
## T-test p-value..... < 2.22e-16      0.32261
## KS Bootstrap p-value.. < 2.22e-16      < 2.22e-16
## KS Naive p-value..... < 2.22e-16      < 2.22e-16
## KS Statistic.....      0.34274      0.22647
##
##
## ***** (V3) education *****
##                               Before Matching      After Matching
## mean treatment.....      10.346      10.346
## mean control.....      12.028      10.328
## std mean diff.....      -83.633      0.86909
##
## mean raw eQQ diff.....      1.7351      0.96962
## med   raw eQQ diff.....      2      1
## max   raw eQQ diff.....      4      6
##
## mean eCDF diff.....      0.090791      0.051033
## med   eCDF diff.....      0.037581      0.024953
## max   eCDF diff.....      0.41227      0.19854
##
## var ratio (Tr/Co).....      0.49052      1.1996
## T-test p-value..... < 2.22e-16      0.92282
## KS Bootstrap p-value.. < 2.22e-16      < 2.22e-16
## KS Naive p-value..... < 2.22e-16      < 2.22e-16
## KS Statistic.....      0.41227      0.19854
##
##
## ***** (V4) I(education^2) *****
##                               Before Matching      After Matching
## mean treatment.....      111.06      111.06
## mean control.....      152.9      110.03
## std mean diff.....      -106.46      2.6216
##
## mean raw eQQ diff.....      42.168      21.122
## med   raw eQQ diff.....      40      21
## max   raw eQQ diff.....      128      128
##
## mean eCDF diff.....      0.090791      0.051033
## med   eCDF diff.....      0.037581      0.024953
## max   eCDF diff.....      0.41227      0.19854

```

```

##
## var ratio (Tr/Co).....    0.34243          1.1728
## T-test p-value..... < 2.22e-16          0.76613
## KS Bootstrap p-value.. < 2.22e-16        < 2.22e-16
## KS Naive p-value..... < 2.22e-16        < 2.22e-16
## KS Statistic.....    0.41227          0.19854
##
##
## ***** (V5) black *****
##                               Before Matching    After Matching
## mean treatment.....    0.84324          0.84324
## mean control.....    0.073537          0.85757
## std mean diff.....    211.13          -3.9306
##
## mean raw eQQ diff.....    0.76757          0.039599
## med  raw eQQ diff.....    1          0
## max  raw eQQ diff.....    1          1
##
## mean eCDF diff.....    0.38485          0.019799
## med  eCDF diff.....    0.38485          0.019799
## max  eCDF diff.....    0.76971          0.039599
##
## var ratio (Tr/Co).....    1.9506          1.0822
## T-test p-value..... < 2.22e-16          0.57579
##
##
## ***** (V6) hispanic *****
##                               Before Matching    After Matching
## mean treatment.....    0.059459          0.059459
## mean control.....    0.072036          0.051035
## std mean diff.....    -5.3038          3.5529
##
## mean raw eQQ diff.....    0.016216          0.032818
## med  raw eQQ diff.....    0          0
## max  raw eQQ diff.....    1          1
##
## mean eCDF diff.....    0.0062883          0.016409
## med  eCDF diff.....    0.0062883          0.016409
## max  eCDF diff.....    0.012577          0.032818
##
## var ratio (Tr/Co).....    0.84109          1.1547
## T-test p-value.....    0.47458          0.71628
##
##
## ***** (V7) married *****
##                               Before Matching    After Matching
## mean treatment.....    0.18919          0.18919
## mean control.....    0.71173          0.20372
## std mean diff.....    -133.06          -3.6997
##

```

```

## mean raw eQQ diff..... 0.51892          0.041768
## med  raw eQQ diff.....      1              0
## max  raw eQQ diff.....      1              1
##
## mean eCDF diff..... 0.26127          0.020884
## med  eCDF diff..... 0.26127          0.020884
## max  eCDF diff..... 0.52254          0.041768
##
## var ratio (Tr/Co)..... 0.75167          0.94562
## T-test p-value..... < 2.22e-16          0.70853
##
##
## ***** (V8) nodegree *****
##                               Before Matching      After Matching
## mean treatment..... 0.70811          0.70811
## mean control..... 0.29584          0.72511
## std mean diff..... 90.437          -3.729
##
## mean raw eQQ diff..... 0.41081          0.079468
## med  raw eQQ diff.....      0              0
## max  raw eQQ diff.....      1              1
##
## mean eCDF diff..... 0.20614          0.039734
## med  eCDF diff..... 0.20614          0.039734
## max  eCDF diff..... 0.41227          0.079468
##
## var ratio (Tr/Co)..... 0.99753          1.0369
## T-test p-value..... < 2.22e-16          0.68656
##
##
## ***** (V9) re74 *****
##                               Before Matching      After Matching
## mean treatment..... 2095.6          2095.6
## mean control..... 14017          2318.3
## std mean diff..... -243.96          -4.558
##
## mean raw eQQ diff..... 12014          3107.6
## med  raw eQQ diff..... 13276          3055.9
## max  raw eQQ diff..... 23256          11202
##
## mean eCDF diff..... 0.45911          0.13406
## med  eCDF diff..... 0.50015          0.12897
## max  eCDF diff..... 0.60309          0.26363
##
## var ratio (Tr/Co)..... 0.26074          0.90315
## T-test p-value..... < 2.22e-16          0.64823
## KS Bootstrap p-value.. < 2.22e-16          < 2.22e-16
## KS Naive p-value..... < 2.22e-16          < 2.22e-16
## KS Statistic..... 0.60309          0.26363
##

```

```

##
## ***** (V10) I(re74^2) *****
##               Before Matching           After Matching
## mean treatment..... 28141412          28141412
## mean control..... 288045960          31671485
## std mean diff..... -227.78           -3.0938
##
## mean raw eQQ diff..... 266205831          70087550
## med  raw eQQ diff..... 221783269          26843876
## max  raw eQQ diff..... 658838222          558946916
##
## mean eCDF diff..... 0.45911           0.13406
## med  eCDF diff..... 0.50015           0.12897
## max  eCDF diff..... 0.60309           0.26363
##
## var ratio (Tr/Co)..... 0.19219          1.1522
## T-test p-value..... < 2.22e-16          0.75399
## KS Bootstrap p-value.. < 2.22e-16        < 2.22e-16
## KS Naive p-value..... < 2.22e-16        < 2.22e-16
## KS Statistic..... 0.60309           0.26363
##
##
## ***** (V11) re75 *****
##               Before Matching           After Matching
## mean treatment..... 1532.1             1532.1
## mean control..... 13651                1688.1
## std mean diff..... -376.45            -4.8483
##
## mean raw eQQ diff..... 12112            1673.1
## med  raw eQQ diff..... 13837            1136.6
## max  raw eQQ diff..... 22438            9715
##
## mean eCDF diff..... 0.4751             0.071313
## med  eCDF diff..... 0.51248            0.069433
## max  eCDF diff..... 0.6509             0.16165
##
## var ratio (Tr/Co)..... 0.12059          0.825
## T-test p-value..... < 2.22e-16          0.62238
## KS Bootstrap p-value.. < 2.22e-16        < 2.22e-16
## KS Naive p-value..... < 2.22e-16        < 2.22e-16
## KS Statistic..... 0.6509             0.16165
##
##
## ***** (V12) I(re75^2) *****
##               Before Matching           After Matching
## mean treatment..... 12654750           12654750
## mean control..... 272279442           15343761
## std mean diff..... -463.34            -4.7989
##
## mean raw eQQ diff..... 259843916        39024771

```



```
## med   raw eQQ diff..... 206883640          11345024
## max   raw eQQ diff..... 629191089          394133437
##
## mean  eCDF diff.....    0.4751          0.071313
## med   eCDF diff.....    0.51248         0.069433
## max   eCDF diff.....    0.6509          0.16165
##
## var ratio (Tr/Co)..... 0.051503          1.0706
## T-test p-value..... < 2.22e-16          0.61373
## KS Bootstrap p-value.. < 2.22e-16        < 2.22e-16
## KS Naive p-value..... < 2.22e-16        < 2.22e-16
## KS Statistic.....    0.6509          0.16165
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): age I(age^2) education I(education^2) black married nodegree re74 I(re74^2) re75 I(re75^2) Number(s): 1 2 3 4 5 7 8 9 10 11 12
##
## After Matching Minimum p.value: < 2.22e-16
## Variable Name(s): age I(age^2) education I(education^2) re74 I(re74^2) re75 I(re75^2) Number(s): 1 2 3 4 9 10 11 12
```

Before Matching Minimum p.value: < 2.22e-16 Variable Name(s): age I(age^2) education I(education^2) black married nodegree re74 I(re74^2) re75 I(re75^2)

After Matching Minimum p.value: < 2.22e-16 Variable Name(s): age I(age^2) education I(education^2) re74 I(re74^2) re75 I(re75^2)

We didn't increase our smallest p-value, but we did reduce the number of variables with small p-values (we improved the p-values for black, married, and nodegree). This Matching increased many of p-values, and also the var ratio (Tr/Co) to be more balanced and closer to 1. This is because propensity score matching only considers one-dimensional space of the "propensity score" to match on; so we lose the valuable information of WHICH covariates affected the "chance to be in the treatment group"; and we don't match units who are actually similar in many parameters, but only in the sense of "how likely they are to be in the treatment group".

## 4: MULTIVARIATE MATCHING.

running a multivariate matching procedure that uses all the covariates and also includes the estimated propensity scores, to produce an estimated treatment effect and confidence intervals.

```
# Carrying out MULTIVARIATE MATCHING (including propensity scores)
cpscombined <- rbind(treated,cps_controls)

# convert {1/0}s to TRUE and FALSE, necessary for functions:
cpscombined$treat <- cpscombined$treat == 1
attach(cpscombined)
```

```

## The following objects are masked from cpscombined (pos = 3):
##
##     age, black, data_id, education, hispanic, married, nodegree,
##     re74, re75, re78, treat

# Estimate the propensity model: Fitting Generalized Linear Models to Later f
ind weights
glm.cps2 <- glm(formula = treat ~ age + I(age^2) + education + I(education^2
) +
               black + hispanic + married + nodegree + re74 + I(re74^2)
               + re75 + I(re75^2) , family=binomial, data=cpscombined)

#summary(glm.cps)
propensities2 = glm.cps$fitted.values

#combining the dataset with the propensity scores
cps.prop <- cbind(cpscombined,propensities)
attach(cps.prop)

## The following object is masked _by_ .GlobalEnv:
##
##     propensities
##
## The following objects are masked from cpscombined (pos = 3):
##
##     age, black, data_id, education, hispanic, married, nodegree,
##     re74, re75, re78, treat

## The following objects are masked from cpscombined (pos = 4):
##
##     age, black, data_id, education, hispanic, married, nodegree,
##     re74, re75, re78, treat

#save data objects for matching according to glm cps
X2.sq <- cbind(propensities2, age , I(age^2) , education , I(education^2),
black , hispanic , married , nodegree , re74 , I(re74^2) ,re75 , I(re75^2))
Y2 <- cps.prop$re78
Tr2 <- cps.prop$treat

# one-to-one matching with replacement (the "M=1" option).
# Estimating the treatment effect on the treated (the "estimand" option defau
lts to ATT).
#
rr2.sq <- Match(Y=Y2, Tr=Tr2, X=X2.sq, M=1, estimand = "ATT", BiasAdj=FALSE,
replace=TRUE, caliper = 0.75, sample = TRUE) ; summary(rr2.sq)

##
## Estimate... 1479
## AI SE..... 825.62
## T-stat..... 1.7914

```

```
## p.val..... 0.073234
##
## Original number of observations..... 16177
## Original number of treated obs..... 185
## Matched number of observations..... 169
## Matched number of observations (unweighted). 205
##
## Caliper (SDs)..... 0.75 0.75 0.75 0.7
5 0.75 0.75 0.75 0.75 0.75 0.75 0.75 0.75 0.75
## Number of obs dropped by 'exact' or 'caliper' 16

#Interpretation: AI SE is matching corrected standard error due to Abadie and Imbens (hence the name AI).
```

## 4.1 Multivariate Matching including second order covariates

Estimate... 1479 AI SE..... 825.62 T-stat..... 1.7914 p.val..... 0.073234 For this matching, since it considers matching onto so many covariates, when the caliper is set to low, it drops many observations without finding suitable counterfactuals for them. Therefore I had to increase the caliper by much. By increasing it to 0.5, the number of dropped obs decreased to 40, and stayed around that until increasing to caliper=0.75 where that number of obs dropped by 'caliper' was **16**.

### (4.1) Finding Confidence Intervals for the estimated treatment effect

*#estimated treatment effect*

```
est2 = rr2.sq$est ; est2
```

```
##          [,1]
```

```
## [1,] 1478.992
```

*#confidence intervals*

```
error = 1.96*(825) ; error
```

```
## [1] 1617
```

```
confint2.low <- est2 - error ; confint2.up <- est2 + error ;
```

```
confint2.low ; confint2.up
```

```
##          [,1]
```

```
## [1,] -138.0078
```

```
##          [,1]
```

```
## [1,] 3095.992
```

### (4.1) Results of Multivariate Matching including second order terms

Here, after 1-1 matching with replacement of caliper 0.75 (distance from all the covariates, including squared ones), the treatment affect mean appears to be a positive **1617**, but the confidence interval {-138,3096} crosses 0 to also the negative side, meaning it's not statistically significant. We can't say with 95% certainty that the mean treatment effect is actually positive at all.

#### (4.1) Checking the covariate balance

*# Checking the covariate balance*

*# 'nboots' is automatically set to small values in the interest of speed, and should be increased to at least 500 each for publication quality p-values.*

```
mb2 <- MatchBalance(treat ~ age + I(age^2) + education + I(education^2) + black +
```

```
  hispanic + married + nodegree + re74 + I(re74^2) + re75  
+ I(re75^2), data=cpscombined, match.out=rr2.sq, nboots=500)
```

```
##
```

```
## ***** (V1) age *****
```

	Before Matching	After Matching
## mean treatment.....	25.816	25.231
## mean control.....	33.225	25.018
## std mean diff.....	-103.55	3.0993
##		
## mean raw eQQ diff.....	7.4865	0.48293
## med raw eQQ diff.....	6	0
## max raw eQQ diff.....	16	3
##		
## mean eCDF diff.....	0.18628	0.015106
## med eCDF diff.....	0.19692	0.0097561
## max eCDF diff.....	0.34274	0.053659
##		
## var ratio (Tr/Co).....	0.41964	0.97867
## T-test p-value..... < 2.22e-16		0.21926
## KS Bootstrap p-value.. < 2.22e-16		0.78
## KS Naive p-value..... < 2.22e-16		0.92943
## KS Statistic.....	0.34274	0.053659

```
##
```

```
##
```

```
## ***** (V2) I(age^2) *****
```

	Before Matching	After Matching
## mean treatment.....	717.39	683.55
## mean control.....	1225.9	673.87
## std mean diff.....	-117.92	2.3518
##		
## mean raw eQQ diff.....	513.91	26.698
## med raw eQQ diff.....	336	0
## max raw eQQ diff.....	1312	297
##		
## mean eCDF diff.....	0.18628	0.015106
## med eCDF diff.....	0.19692	0.0097561
## max eCDF diff.....	0.34274	0.053659
##		
## var ratio (Tr/Co).....	0.302	1.0263
## T-test p-value..... < 2.22e-16		0.32866
## KS Bootstrap p-value.. < 2.22e-16		0.78
## KS Naive p-value..... < 2.22e-16		0.92943
## KS Statistic.....	0.34274	0.053659

```

##
##
## ***** (V3) education *****
##               Before Matching      After Matching
## mean treatment.....      10.346      10.473
## mean control.....      12.028      10.538
## std mean diff.....      -83.633      -3.671
##
## mean raw eQQ diff.....      1.7351      0.063415
## med  raw eQQ diff.....      2      0
## max  raw eQQ diff.....      4      1
##
## mean eCDF diff.....      0.090791      0.004878
## med  eCDF diff.....      0.037581      0
## max  eCDF diff.....      0.41227      0.029268
##
## var ratio (Tr/Co).....      0.49052      1.0912
## T-test p-value..... < 2.22e-16      0.24337
## KS Bootstrap p-value.. < 2.22e-16      0.95
## KS Naive p-value..... < 2.22e-16      0.99999
## KS Statistic.....      0.41227      0.029268
##
##
## ***** (V4) I(education^2) *****
##               Before Matching      After Matching
## mean treatment.....      111.06      112.82
## mean control.....      152.9      113.92
## std mean diff.....      -106.46      -3.0314
##
## mean raw eQQ diff.....      42.168      1.1171
## med  raw eQQ diff.....      40      0
## max  raw eQQ diff.....      128      21
##
## mean eCDF diff.....      0.090791      0.004878
## med  eCDF diff.....      0.037581      0
## max  eCDF diff.....      0.41227      0.029268
##
## var ratio (Tr/Co).....      0.34243      1.0676
## T-test p-value..... < 2.22e-16      0.30323
## KS Bootstrap p-value.. < 2.22e-16      0.95
## KS Naive p-value..... < 2.22e-16      0.99999
## KS Statistic.....      0.41227      0.029268
##
##
## ***** (V5) black *****
##               Before Matching      After Matching
## mean treatment.....      0.84324      0.84024
## mean control.....      0.073537      0.84024
## std mean diff.....      211.13      0
##

```

```

## mean raw eQQ diff..... 0.76757 0
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.38485 0
## med eCDF diff..... 0.38485 0
## max eCDF diff..... 0.76971 0
##
## var ratio (Tr/Co)..... 1.9506 1
## T-test p-value..... < 2.22e-16 1
##
##
## ***** (V6) hispanic *****
## Before Matching After Matching
## mean treatment..... 0.059459 0.053254
## mean control..... 0.072036 0.053254
## std mean diff..... -5.3038 0
##
## mean raw eQQ diff..... 0.016216 0
## med raw eQQ diff..... 0 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.0062883 0
## med eCDF diff..... 0.0062883 0
## max eCDF diff..... 0.012577 0
##
## var ratio (Tr/Co)..... 0.84109 1
## T-test p-value..... 0.47458 1
##
##
## ***** (V7) married *****
## Before Matching After Matching
## mean treatment..... 0.18919 0.18935
## mean control..... 0.71173 0.18935
## std mean diff..... -133.06 0
##
## mean raw eQQ diff..... 0.51892 0
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.26127 0
## med eCDF diff..... 0.26127 0
## max eCDF diff..... 0.52254 0
##
## var ratio (Tr/Co)..... 0.75167 1
## T-test p-value..... < 2.22e-16 1
##
##
## ***** (V8) nodegree *****
## Before Matching After Matching

```

```

## mean treatment..... 0.70811          0.69822
## mean control..... 0.29584          0.69822
## std mean diff..... 90.437           0
##
## mean raw eQQ diff..... 0.41081          0
## med raw eQQ diff..... 0              0
## max raw eQQ diff..... 1              0
##
## mean eCDF diff..... 0.20614          0
## med eCDF diff..... 0.20614          0
## max eCDF diff..... 0.41227          0
##
## var ratio (Tr/Co)..... 0.99753          1
## T-test p-value..... < 2.22e-16        1
##
##
## ***** (V9) re74 *****
##                               Before Matching    After Matching
## mean treatment..... 2095.6          1722.8
## mean control..... 14017            1720.6
## std mean diff..... -243.96         0.058784
##
## mean raw eQQ diff..... 12014         189.42
## med raw eQQ diff..... 13276          0
## max raw eQQ diff..... 23256         2828.9
##
## mean eCDF diff..... 0.45911         0.01694
## med eCDF diff..... 0.50015         0.0097561
## max eCDF diff..... 0.60309         0.097561
##
## var ratio (Tr/Co)..... 0.26074         1.1141
## T-test p-value..... < 2.22e-16        0.98557
## KS Bootstrap p-value.. < 2.22e-16        0.102
## KS Naive p-value..... < 2.22e-16        0.28341
## KS Statistic..... 0.60309         0.097561
##
##
## ***** (V10) I(re74^2) *****
##                               Before Matching    After Matching
## mean treatment..... 28141412         16877544
## mean control..... 288045960         15445332
## std mean diff..... -227.78         2.7827
##
## mean raw eQQ diff..... 266205831         2492622
## med raw eQQ diff..... 221783269          0
## max raw eQQ diff..... 658838222         64700233
##
## mean eCDF diff..... 0.45911         0.01694
## med eCDF diff..... 0.50015         0.0097561
## max eCDF diff..... 0.60309         0.097561

```

```

##
## var ratio (Tr/Co).....    0.19219          1.0104
## T-test p-value..... < 2.22e-16          0.32779
## KS Bootstrap p-value.. < 2.22e-16          0.102
## KS Naive p-value..... < 2.22e-16          0.28341
## KS Statistic.....    0.60309          0.097561
##
##
## ***** (V11) re75 *****
##                               Before Matching    After Matching
## mean treatment.....    1532.1          1503.4
## mean control.....    13651          1440.2
## std mean diff.....    -376.45          1.9708
##
## mean raw eQQ diff.....    12112          167.16
## med  raw eQQ diff.....    13837          0
## max  raw eQQ diff.....    22438          3195
##
## mean eCDF diff.....    0.4751          0.018321
## med  eCDF diff.....    0.51248          0.014634
## max  eCDF diff.....    0.6509          0.058537
##
## var ratio (Tr/Co).....    0.12059          0.96875
## T-test p-value..... < 2.22e-16          0.57282
## KS Bootstrap p-value.. < 2.22e-16          0.578
## KS Naive p-value..... < 2.22e-16          0.87388
## KS Statistic.....    0.6509          0.058537
##
##
## ***** (V12) I(re75^2) *****
##                               Before Matching    After Matching
## mean treatment.....    12654750          12503600
## mean control.....    272279442          12647762
## std mean diff.....    -463.34          -0.24972
##
## mean raw eQQ diff.....    259843916          1614383
## med  raw eQQ diff.....    206883640          0
## max  raw eQQ diff.....    629191089          80124381
##
## mean eCDF diff.....    0.4751          0.018321
## med  eCDF diff.....    0.51248          0.014634
## max  eCDF diff.....    0.6509          0.058537
##
## var ratio (Tr/Co).....    0.051503          0.94534
## T-test p-value..... < 2.22e-16          0.87336
## KS Bootstrap p-value.. < 2.22e-16          0.578
## KS Naive p-value..... < 2.22e-16          0.87388
## KS Statistic.....    0.6509          0.058537
##
##
##

```



```
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): age I(age^2) education I(education^2) black married nodegree re74 I(re74^2) re75 I(re75^2) Number(s): 1 2 3 4 5 7 8 9 10 11 12
##
## After Matching Minimum p.value: 0.102
## Variable Name(s): re74 I(re74^2) Number(s): 9 10
```

We didn't increase our smallest p-value, but we did reduce the number of variables with small p-values and increased the majority of p-values to be much higher.

## 4.2 Multivariate Matching of only Linear terms

To see if our matching balance improves if we reduce the number of necessary covariates - the second order terms, we'll check that too:

```
#defining another set of predictors without the squared terms, for a second trial:
X2.lin <- cbind(propensities, age , education ,
               black , hispanic , married , nodegree , re74 ,
               re75)

#multivariate matching with only linear, first order terms:
rr2.lin.caliper <- Match(Y=Y2, Tr=Tr2, X=X2.lin, M=1, estimand = "ATT", BiasAdj=FALSE, replace=TRUE, caliper = 0.75, sample = TRUE) ; summary(rr2.lin.caliper)

##
## Estimate... 1972.5
## AI SE..... 920.67
## T-stat..... 2.1425
## p.val..... 0.032155
##
## Original number of observations..... 16177
## Original number of treated obs..... 185
## Matched number of observations..... 171
## Matched number of observations (unweighted). 208
##
## Caliper (SDs)..... 0.75 0.75 0.75 0.75
5 0.75 0.75 0.75 0.75 0.75
## Number of obs dropped by 'exact' or 'caliper' 14

rr2.lin.nocaliper <- Match(Y=Y2, Tr=Tr2, X=X2.lin, M=1, estimand = "ATT", BiasAdj=FALSE, replace=TRUE, sample = TRUE) ; summary(rr2.lin.nocaliper)

##
## Estimate... 2071
## AI SE..... 966.97
## T-stat..... 2.1417
## p.val..... 0.032218
##
## Original number of observations..... 16177
```



```

## mean treatment..... 0.37962          0.37074
## mean control..... 0.0071768          0.36763
## std mean diff..... 135.98           1.1189
##
## mean raw eQQ diff..... 0.36946          0.0070305
## med raw eQQ diff..... 0.35183          0.0040108
## max raw eQQ diff..... 0.83504          0.035677
##
## mean eCDF diff..... 0.49836          0.0092935
## med eCDF diff..... 0.53052          0.0096154
## max eCDF diff..... 0.87124          0.038462
##
## var ratio (Tr/Co)..... 41.251          0.98711
## T-test p-value..... < 2.22e-16          0.013752
## KS Bootstrap p-value.. < 2.22e-16          0.992
## KS Naive p-value..... < 2.22e-16          0.9979
## KS Statistic..... 0.87124          0.038462
##
##
## ***** (V2) age *****
##                               Before Matching      After Matching
## mean treatment..... 25.816          25.421
## mean control..... 33.225          25.205
## std mean diff..... -103.55          3.0673
##
## mean raw eQQ diff..... 7.4865          0.57212
## med raw eQQ diff..... 6          1
## max raw eQQ diff..... 16          3
##
## mean eCDF diff..... 0.18628          0.01799
## med eCDF diff..... 0.19692          0.0096154
## max eCDF diff..... 0.34274          0.067308
##
## var ratio (Tr/Co)..... 0.41964          0.98119
## T-test p-value..... < 2.22e-16          0.22042
## KS Bootstrap p-value.. < 2.22e-16          0.568
## KS Naive p-value..... < 2.22e-16          0.73373
## KS Statistic..... 0.34274          0.067308
##
##
## ***** (V3) education *****
##                               Before Matching      After Matching
## mean treatment..... 10.346          10.456
## mean control..... 12.028          10.52
## std mean diff..... -83.633          -3.4754
##
## mean raw eQQ diff..... 1.7351          0.10096
## med raw eQQ diff..... 2          0
## max raw eQQ diff..... 4          2
##

```

```

## mean eCDF diff..... 0.090791      0.0068681
## med eCDF diff..... 0.037581      0.0048077
## max eCDF diff..... 0.41227      0.024038
##
## var ratio (Tr/Co)..... 0.49052      1.0206
## T-test p-value..... < 2.22e-16      0.29639
## KS Bootstrap p-value.. < 2.22e-16      0.976
## KS Naive p-value..... < 2.22e-16      1
## KS Statistic..... 0.41227      0.024038
##
##
## ***** (V4) black *****
##
## Before Matching      After Matching
## mean treatment..... 0.84324      0.84211
## mean control..... 0.073537      0.84211
## std mean diff..... 211.13      0
##
## mean raw eQQ diff..... 0.76757      0
## med raw eQQ diff..... 1      0
## max raw eQQ diff..... 1      0
##
## mean eCDF diff..... 0.38485      0
## med eCDF diff..... 0.38485      0
## max eCDF diff..... 0.76971      0
##
## var ratio (Tr/Co)..... 1.9506      1
## T-test p-value..... < 2.22e-16      1
##
##
## ***** (V5) hispanic *****
##
## Before Matching      After Matching
## mean treatment..... 0.059459      0.052632
## mean control..... 0.072036      0.052632
## std mean diff..... -5.3038      0
##
## mean raw eQQ diff..... 0.016216      0
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 1      0
##
## mean eCDF diff..... 0.0062883      0
## med eCDF diff..... 0.0062883      0
## max eCDF diff..... 0.012577      0
##
## var ratio (Tr/Co)..... 0.84109      1
## T-test p-value..... 0.47458      1
##
##
## ***** (V6) married *****
##
## Before Matching      After Matching
## mean treatment..... 0.18919      0.19298

```

```

## mean control..... 0.71173 0.19298
## std mean diff..... -133.06 0
##
## mean raw eQQ diff..... 0.51892 0
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.26127 0
## med eCDF diff..... 0.26127 0
## max eCDF diff..... 0.52254 0
##
## var ratio (Tr/Co)..... 0.75167 1
## T-test p-value..... < 2.22e-16 1
##
##
## ***** (V7) nodegree *****
## Before Matching After Matching
## mean treatment..... 0.70811 0.69591
## mean control..... 0.29584 0.69591
## std mean diff..... 90.437 0
##
## mean raw eQQ diff..... 0.41081 0
## med raw eQQ diff..... 0 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.20614 0
## med eCDF diff..... 0.20614 0
## max eCDF diff..... 0.41227 0
##
## var ratio (Tr/Co)..... 0.99753 1
## T-test p-value..... < 2.22e-16 1
##
##
## ***** (V8) re74 *****
## Before Matching After Matching
## mean treatment..... 2095.6 1702.7
## mean control..... 14017 1704.7
## std mean diff..... -243.96 -0.055188
##
## mean raw eQQ diff..... 12014 142.38
## med raw eQQ diff..... 13276 0
## max raw eQQ diff..... 23256 2271.3
##
## mean eCDF diff..... 0.45911 0.011953
## med eCDF diff..... 0.50015 0.0096154
## max eCDF diff..... 0.60309 0.086538
##
## var ratio (Tr/Co)..... 0.26074 1.0324
## T-test p-value..... < 2.22e-16 0.98561
## KS Bootstrap p-value.. < 2.22e-16 0.124

```

```
## KS Naive p-value..... < 2.22e-16          0.41731
## KS Statistic.....      0.60309          0.086538
##
##
## ***** (V9) re75 *****
##               Before Matching      After Matching
## mean treatment.....      1532.1      1485.8
## mean control.....      13651      1365.2
## std mean diff.....     -376.45       3.776
##
## mean raw eQQ diff.....      12112      232.99
## med  raw eQQ diff.....      13837       0
## max  raw eQQ diff.....      22438     5189.1
##
## mean eCDF diff.....      0.4751     0.028115
## med  eCDF diff.....      0.51248     0.024038
## max  eCDF diff.....      0.6509     0.072115
##
## var ratio (Tr/Co).....      0.12059      1.0854
## T-test p-value..... < 2.22e-16      0.28528
## KS Bootstrap p-value.. < 2.22e-16      0.348
## KS Naive p-value..... < 2.22e-16      0.65172
## KS Statistic.....      0.6509     0.072115
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): cps.prop$propensities age education black married nodegr
ee re74 re75  Number(s): 1 2 3 4 6 7 8 9
##
## After Matching Minimum p.value: 0.013752
## Variable Name(s): cps.prop$propensities  Number(s): 1
```

**\*\*RESULTS:** Before Matching Minimum p.value: < 2.22e-16 Variable Name(s):  
cps.prop\$propensities age education black married nodegree re74 re75 Number(s): 1 2 3 4  
6 7 8 9

After Matching Minimum p.value: 0.013752 Variable Name(s): cps.prop\$propensities  
Number(s): 1

With the nonlinear terms, we actually got a slightly better lowest p value: 0.07 vs. p=0.013.  
However, with no linear terms, this low p-value was found only for 1 variable, whereas for  
the

## 5: GenMatch

### 5.1: GenMatch Propensity Scores Only

```
#The covariates we want to match on
X51 <- propensities
```

```

#The covariates we want to obtain balance on
BalanceMat51 <- propensities

#Let's call GenMatch() to find the optimal weight to give each
#covariate in 'X' so as we have achieved balance on the covariates in #'BalanceMat'

genout <- GenMatch(Tr=treat, X=X51, BalanceMatrix=BalanceMat51, estimand="ATT",
M=1, pop.size=100, max.generations=50, replace = TRUE, wait.generations=4
) #I increased the pop.size to 150 and max.generations to 100 but not for this run

##
##
## Wed Mar 22 22:48:13 2017
## Domains:
## 0.000000e+00 <= X1 <= 1.000000e+03
##
## Data Type: Floating Point
## Operators (code number, name, population)
## (1) Cloning..... 15
## (2) Uniform Mutation..... 12
## (3) Boundary Mutation..... 12
## (4) Non-Uniform Mutation..... 12
## (5) Polytope Crossover..... 12
## (6) Simple Crossover..... 12
## (7) Whole Non-Uniform Mutation..... 12
## (8) Heuristic Crossover..... 12
## (9) Local-Minimum Crossover..... 0
##
## SOFT Maximum Number of Generations: 50
## Maximum Nonchanging Generations: 4
## Population size : 100
## Convergence Tolerance: 1.000000e-03
##
## Not Using the BFGS Derivative Based Optimizer on the Best Individual Each
Generation.
## Not Checking Gradients before Stopping.
## Using Out of Bounds Individuals.
##
## Maximization Problem.
## GENERATION: 0 (initializing the population)
## Lexical Fit..... 7.179235e-01 8.111591e-01
## #unique..... 100, #Total UniqueCount: 100
## var 1:
## best..... 5.950875e+02
## mean..... 5.173951e+02
## variance..... 9.493123e+04
##
## GENERATION: 1

```

```
## Lexical Fit..... 7.179235e-01  8.111591e-01
## #unique..... 63, #Total UniqueCount: 163
## var 1:
## best..... 5.950875e+02
## mean..... 5.957691e+02
## variance..... 2.914700e+04
##
## GENERATION: 2
## Lexical Fit..... 7.179235e-01  8.111591e-01
## #unique..... 58, #Total UniqueCount: 221
## var 1:
## best..... 5.950875e+02
## mean..... 5.944606e+02
## variance..... 1.778970e+04
##
## GENERATION: 3
## Lexical Fit..... 7.179235e-01  8.111591e-01
## #unique..... 57, #Total UniqueCount: 278
## var 1:
## best..... 5.950875e+02
## mean..... 5.796034e+02
## variance..... 1.056714e+04
##
## GENERATION: 4
## Lexical Fit..... 7.179235e-01  8.111591e-01
## #unique..... 56, #Total UniqueCount: 334
## var 1:
## best..... 5.950875e+02
## mean..... 5.941342e+02
## variance..... 1.226961e+04
##
## GENERATION: 5
## Lexical Fit..... 7.179235e-01  8.111591e-01
## #unique..... 56, #Total UniqueCount: 390
## var 1:
## best..... 5.950875e+02
## mean..... 5.812610e+02
## variance..... 9.328822e+03
##
## 'wait.generations' limit reached.
## No significant improvement in 4 generations.
##
## Solution Lexical Fitness Value:
## 7.179235e-01  8.111591e-01
##
## Parameters at the Solution:
##
## X[ 1] : 5.950875e+02
##
## Solution Found Generation 1
```



```

## Number of Generations Run 5
##
## Wed Mar 22 22:48:53 2017
## Total run time : 0 hours 0 minutes and 40 seconds

#genout

#The outcome variable
Y=re78

#
# Now that GenMatch() has found the optimal weights, let's estimate
# our causal effect of interest using those weights (using Weight.matrix=genout)

mout1 <- Match(Y=Y, Tr=treat, X=X51, estimand="ATT", M=1, replace = TRUE, Weight.matrix=genout)
summary(mout1)

##
## Estimate... 1937.9
## AI SE..... 993.39
## T-stat..... 1.9507
## p.val..... 0.051087
##
## Original number of observations..... 16177
## Original number of treated obs..... 185
## Matched number of observations..... 185
## Matched number of observations (unweighted). 315

#
#Let's determine if balance has actually been obtained on the variables of interest
#
mb1 <- MatchBalance(treat~ propensities+age+education+black+hispanic+ married
+ nodegree+ re75+ re74, match.out=mout1, nboots=500)

##
## ***** (V1) propensities *****
##
##          Before Matching      After Matching
## mean treatment..... 0.37962      0.37962
## mean control..... 0.0071768      0.37973
## std mean diff..... 135.98      -0.043488
##
## mean raw eQQ diff..... 0.36946      0.001436
## med raw eQQ diff..... 0.35183      9.0877e-06
## max raw eQQ diff..... 0.83504      0.019616
##
## mean eCDF diff..... 0.49836      0.0059405
## med eCDF diff..... 0.53052      0.0031746

```

```

## max eCDF diff..... 0.87124 0.050794
##
## var ratio (Tr/Co)..... 41.251 0.99792
## T-test p-value..... < 2.22e-16 0.71792
## KS Bootstrap p-value.. < 2.22e-16 0.8
## KS Naive p-value..... < 2.22e-16 0.81116
## KS Statistic..... 0.87124 0.050794
##
##
## ***** (V2) age *****
## Before Matching After Matching
## mean treatment..... 25.816 25.816
## mean control..... 33.225 26.407
## std mean diff..... -103.55 -8.2622
##
## mean raw eQQ diff..... 7.4865 1.7873
## med raw eQQ diff..... 6 1
## max raw eQQ diff..... 16 5
##
## mean eCDF diff..... 0.18628 0.055853
## med eCDF diff..... 0.19692 0.053968
## max eCDF diff..... 0.34274 0.13968
##
## var ratio (Tr/Co)..... 0.41964 0.89245
## T-test p-value..... < 2.22e-16 0.3883
## KS Bootstrap p-value.. < 2.22e-16 0.002
## KS Naive p-value..... < 2.22e-16 0.0042839
## KS Statistic..... 0.34274 0.13968
##
##
## ***** (V3) education *****
## Before Matching After Matching
## mean treatment..... 10.346 10.346
## mean control..... 12.028 10.302
## std mean diff..... -83.633 2.166
##
## mean raw eQQ diff..... 1.7351 0.31429
## med raw eQQ diff..... 2 0
## max raw eQQ diff..... 4 4
##
## mean eCDF diff..... 0.090791 0.019048
## med eCDF diff..... 0.037581 0.014286
## max eCDF diff..... 0.41227 0.053968
##
## var ratio (Tr/Co)..... 0.49052 1.0992
## T-test p-value..... < 2.22e-16 0.81579
## KS Bootstrap p-value.. < 2.22e-16 0.334
## KS Naive p-value..... < 2.22e-16 0.74862
## KS Statistic..... 0.41227 0.053968
##

```

```

##
## ***** (V4) black *****
##                               Before Matching      After Matching
## mean treatment.....      0.84324      0.84324
## mean control.....      0.073537      0.84935
## std mean diff.....      211.13      -1.6756
##
## mean raw eQQ diff.....      0.76757      0.060317
## med raw eQQ diff.....      1      0
## max raw eQQ diff.....      1      1
##
## mean eCDF diff.....      0.38485      0.030159
## med eCDF diff.....      0.38485      0.030159
## max eCDF diff.....      0.76971      0.060317
##
## var ratio (Tr/Co).....      1.9506      1.0331
## T-test p-value..... < 2.22e-16      0.81178
##
##
## ***** (V5) hispanic *****
##                               Before Matching      After Matching
## mean treatment.....      0.059459      0.059459
## mean control.....      0.072036      0.047062
## std mean diff.....      -5.3038      5.2282
##
## mean raw eQQ diff.....      0.016216      0.019048
## med raw eQQ diff.....      0      0
## max raw eQQ diff.....      1      1
##
## mean eCDF diff.....      0.0062883      0.0095238
## med eCDF diff.....      0.0062883      0.0095238
## max eCDF diff.....      0.012577      0.019048
##
## var ratio (Tr/Co).....      0.84109      1.247
## T-test p-value.....      0.47458      0.58069
##
##
## ***** (V6) married *****
##                               Before Matching      After Matching
## mean treatment.....      0.18919      0.18919
## mean control.....      0.71173      0.19687
## std mean diff.....      -133.06      -1.9565
##
## mean raw eQQ diff.....      0.51892      0.019048
## med raw eQQ diff.....      1      0
## max raw eQQ diff.....      1      1
##
## mean eCDF diff.....      0.26127      0.0095238
## med eCDF diff.....      0.26127      0.0095238
## max eCDF diff.....      0.52254      0.019048

```

```

##
## var ratio (Tr/Co).....    0.75167          0.97017
## T-test p-value..... < 2.22e-16          0.84335
##
##
## ***** (V7) nodegree *****
##                               Before Matching      After Matching
## mean treatment.....    0.70811          0.70811
## mean control.....    0.29584          0.71756
## std mean diff.....    90.437          -2.0743
##
## mean raw eQQ diff.....    0.41081          0.025397
## med  raw eQQ diff.....    0          0
## max  raw eQQ diff.....    1          1
##
## mean eCDF diff.....    0.20614          0.012698
## med  eCDF diff.....    0.20614          0.012698
## max  eCDF diff.....    0.41227          0.025397
##
## var ratio (Tr/Co).....    0.99753          1.0199
## T-test p-value..... < 2.22e-16          0.82111
##
##
## ***** (V8) re75 *****
##                               Before Matching      After Matching
## mean treatment.....    1532.1          1532.1
## mean control.....    13651          1550
## std mean diff.....    -376.45          -0.55757
##
## mean raw eQQ diff.....    12112          1392.1
## med  raw eQQ diff.....    13837          156.5
## max  raw eQQ diff.....    22438          14221
##
## mean eCDF diff.....    0.4751          0.061649
## med  eCDF diff.....    0.51248          0.063492
## max  eCDF diff.....    0.6509          0.10159
##
## var ratio (Tr/Co).....    0.12059          1.0484
## T-test p-value..... < 2.22e-16          0.95228
## KS Bootstrap p-value.. < 2.22e-16          0.052
## KS Naive p-value..... < 2.22e-16          0.077482
## KS Statistic.....    0.6509          0.10159
##
##
## ***** (V9) re74 *****
##                               Before Matching      After Matching
## mean treatment.....    2095.6          2095.6
## mean control.....    14017          2139.5
## std mean diff.....    -243.96          -0.89911
##

```

```
## mean raw eQQ diff.....      12014      1339.2
## med  raw eQQ diff.....      13276      0
## max  raw eQQ diff.....      23256      9177.8
##
## mean eCDF diff.....      0.45911      0.076063
## med  eCDF diff.....      0.50015      0.069841
## max  eCDF diff.....      0.60309      0.15873
##
## var ratio (Tr/Co).....      0.26074      1.0373
## T-test p-value..... < 2.22e-16      0.92517
## KS Bootstrap p-value.. < 2.22e-16      < 2.22e-16
## KS Naive p-value..... < 2.22e-16      0.00071491
## KS Statistic.....      0.60309      0.15873
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): propensities age education black married nodegree re75 re74
## Number(s): 1 2 3 4 6 7 8 9
##
## After Matching Minimum p.value: < 2.22e-16
## Variable Name(s): re74 Number(s): 9
```

### GenMatch for Propensity Scores Matching Results:

Before Matching Minimum p.value: < 2.22e-16 Variable Name(s): age education black married nodegree re75 re74 Number(s): 1 2 3 5 6 7 8

After Matching Minimum p.value: < 2.22e-16 Variable Name(s): age education black married nodegree re75 re74 Number(s): 1 2 3 5 6 7 8

We see practically no improvement in the matching balance. The matching didn't achieve better balance than before: not in the lower p-value, and not in the amount of variables with that low p-value. This is because the `GenMatch()` is meant to get the weights for all the covariates, but if we end up only matching on propensity scores, then it does nothing in practice.

Although this estimate is indeed relatively close to the original experimental treatment effect mentioned in section 1: 1794, it's actually not really closer than previous estimations, for example, in question #3.

The Genetic Matching Function actually only considers ONE COVARIATE, the propensity scores; therefore, it doesn't do its job of optimizing different weights for different parameters; it assigns everything onto that one covariate, and just checks what's the optimal covariate for it. We can see that demonstrated in the resulting weight matrix from the `genmatch`, which contains only one weight!

```
$Weight.matrix[,1][1,] 910.1453
```

### (5.1) Confidence Intervals

```

#estimated treatment effect
est51 = 1937.9 ; est51

## [1] 1937.9

#confidence intervals
error51 = 1.96*(993.39) ; error51

## [1] 1947.044

confint51.low <- est51 - error51 ; confint51.up <- est51 + error51 ;
confint51.low ; confint51.up

## [1] -9.1444

## [1] 3884.944

```

434.52 <--(mean:1890.8)--> 3347.08 Here we see that the treatment effect is positive, with narrower confidence intervals (although not by much than our best results outside of Genetic Matching).

## 5.2 - GenMatch for Multivariate Matching

```

#The covariates we want to match on
X52 = cbind(propensities, age, education, black, hispanic, married, nodegree,
re74, re75)

#The covariates we want to obtain balance on are the same as X52

#Let's call GenMatch() to find the optimal weight to give each
#covariate in 'X' so as we have achieved balance on the covariates in #'BalanceMat'

genout2 <- GenMatch(Tr=treat, X=X52, BalanceMatrix = X52, estimand="ATT", M=1
, pop.size=150, max.generations=150, replace = TRUE, wait.generations=4,caliper = 0.45)

##
##
## Wed Mar 22 22:49:09 2017
## Domains:
## 0.000000e+00 <= X1 <= 1.000000e+03
## 0.000000e+00 <= X2 <= 1.000000e+03
## 0.000000e+00 <= X3 <= 1.000000e+03
## 0.000000e+00 <= X4 <= 1.000000e+03
## 0.000000e+00 <= X5 <= 1.000000e+03
## 0.000000e+00 <= X6 <= 1.000000e+03
## 0.000000e+00 <= X7 <= 1.000000e+03
## 0.000000e+00 <= X8 <= 1.000000e+03
## 0.000000e+00 <= X9 <= 1.000000e+03
##
## Data Type: Floating Point

```

```

## Operators (code number, name, population)
## (1) Cloning..... 14
## (2) Uniform Mutation..... 19
## (3) Boundary Mutation..... 19
## (4) Non-Uniform Mutation..... 19
## (5) Polytope Crossover..... 19
## (6) Simple Crossover..... 20
## (7) Whole Non-Uniform Mutation..... 19
## (8) Heuristic Crossover..... 20
## (9) Local-Minimum Crossover..... 0
##
## SOFT Maximum Number of Generations: 150
## Maximum Nonchanging Generations: 4
## Population size      : 150
## Convergence Tolerance: 1.000000e-03
##
## Not Using the BFGS Derivative Based Optimizer on the Best Individual Each
## Generation.
## Not Checking Gradients before Stopping.
## Using Out of Bounds Individuals.
##
## Maximization Problem.
## GENERATION: 0 (initializing the population)
## Lexical Fit..... 4.352311e-01  6.814275e-01  6.955913e-01  6.992493e-01  7
.165829e-01  8.176879e-01  9.658675e-01  9.658675e-01  9.999992e-01  9.999992
e-01  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1
.000000e+00  1.000000e+00  1.000000e+00
## #unique..... 150, #Total UniqueCount: 150
## var 1:
## best..... 1.753650e+02
## mean..... 5.123283e+02
## variance..... 8.890979e+04
## var 2:
## best..... 5.656405e+02
## mean..... 4.950556e+02
## variance..... 7.938807e+04
## var 3:
## best..... 2.088827e+02
## mean..... 5.012546e+02
## variance..... 8.261712e+04
## var 4:
## best..... 2.485503e+02
## mean..... 4.966375e+02
## variance..... 8.679036e+04
## var 5:
## best..... 2.456207e+02
## mean..... 4.855677e+02
## variance..... 8.345624e+04
## var 6:
## best..... 1.110987e+02

```

```
## mean..... 4.728829e+02
## variance..... 8.656510e+04
## var 7:
## best..... 3.089612e+02
## mean..... 4.967383e+02
## variance..... 9.587667e+04
## var 8:
## best..... 1.745945e+00
## mean..... 4.910211e+02
## variance..... 8.869159e+04
## var 9:
## best..... 1.290457e+02
## mean..... 4.584630e+02
## variance..... 7.691360e+04
##
## GENERATION: 1
## Lexical Fit..... 5.146841e-01 5.558456e-01 5.642060e-01 6.814275e-01 9
.658675e-01 9.879251e-01 9.975381e-01 9.983196e-01 9.999992e-01 1.000000
e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1
.000000e+00 1.000000e+00 1.000000e+00
## #unique..... 114, #Total UniqueCount: 264
## var 1:
## best..... 2.333475e+02
## mean..... 3.172654e+02
## variance..... 4.476369e+04
## var 2:
## best..... 5.587115e+02
## mean..... 4.854771e+02
## variance..... 8.948761e+04
## var 3:
## best..... 2.088827e+02
## mean..... 3.079563e+02
## variance..... 6.904869e+04
## var 4:
## best..... 6.630301e+02
## mean..... 3.905459e+02
## variance..... 4.979435e+04
## var 5:
## best..... 1.775254e+02
## mean..... 4.553481e+02
## variance..... 7.243103e+04
## var 6:
## best..... 1.110987e+02
## mean..... 2.603857e+02
## variance..... 6.712372e+04
## var 7:
## best..... 4.216371e+02
## mean..... 3.626785e+02
## variance..... 4.136272e+04
## var 8:
```



```
## best..... 1.745945e+00
## mean..... 9.779713e+01
## variance..... 3.663084e+04
## var 9:
## best..... 1.290457e+02
## mean..... 4.621053e+02
## variance..... 9.485526e+04
##
## GENERATION: 2
## Lexical Fit..... 5.851030e-01 6.314611e-01 6.814275e-01 8.285723e-01 8
.815982e-01 9.206334e-01 9.658675e-01 9.809493e-01 9.999992e-01 9.999992
e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1
.000000e+00 1.000000e+00 1.000000e+00
## #unique..... 118, #Total UniqueCount: 382
## var 1:
## best..... 5.298648e+02
## mean..... 2.932249e+02
## variance..... 2.708303e+04
## var 2:
## best..... 6.450107e+02
## mean..... 5.362644e+02
## variance..... 1.828035e+04
## var 3:
## best..... 7.204414e+02
## mean..... 3.050617e+02
## variance..... 4.596460e+04
## var 4:
## best..... 6.915332e+02
## mean..... 5.039468e+02
## variance..... 4.799532e+04
## var 5:
## best..... 2.366926e+02
## mean..... 2.765857e+02
## variance..... 2.682408e+04
## var 6:
## best..... 1.305135e+02
## mean..... 1.515566e+02
## variance..... 1.201627e+04
## var 7:
## best..... 2.855952e+02
## mean..... 3.718250e+02
## variance..... 1.434173e+04
## var 8:
## best..... 1.677894e+01
## mean..... 4.557737e+01
## variance..... 1.626329e+04
## var 9:
## best..... 1.969262e+02
## mean..... 2.084327e+02
## variance..... 3.337494e+04
```

```

##
## GENERATION: 3
## Lexical Fit..... 5.851030e-01  6.314611e-01  6.814275e-01  8.285723e-01  8
.815982e-01  9.206334e-01  9.658675e-01  9.809493e-01  9.999992e-01  9.999992
e-01  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1
.000000e+00  1.000000e+00  1.000000e+00
## #unique..... 110, #Total UniqueCount: 492
## var 1:
## best..... 5.298648e+02
## mean..... 4.621885e+02
## variance..... 1.939668e+04
## var 2:
## best..... 6.450107e+02
## mean..... 6.005408e+02
## variance..... 6.951399e+03
## var 3:
## best..... 7.204414e+02
## mean..... 5.966880e+02
## variance..... 4.668005e+04
## var 4:
## best..... 6.915332e+02
## mean..... 6.761487e+02
## variance..... 1.193911e+04
## var 5:
## best..... 2.366926e+02
## mean..... 2.465552e+02
## variance..... 1.124379e+04
## var 6:
## best..... 1.305135e+02
## mean..... 1.742430e+02
## variance..... 1.623773e+04
## var 7:
## best..... 2.855952e+02
## mean..... 3.263885e+02
## variance..... 1.077852e+04
## var 8:
## best..... 1.677894e+01
## mean..... 3.589563e+01
## variance..... 1.017920e+04
## var 9:
## best..... 1.969262e+02
## mean..... 1.794908e+02
## variance..... 5.343483e+03
##
## GENERATION: 4
## Lexical Fit..... 5.851030e-01  6.314611e-01  6.814275e-01  8.285723e-01  8
.815982e-01  9.206334e-01  9.658675e-01  9.809493e-01  9.999992e-01  9.999992
e-01  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1
.000000e+00  1.000000e+00  1.000000e+00
## #unique..... 103, #Total UniqueCount: 595

```

```

## var 1:
## best..... 5.298648e+02
## mean..... 5.154838e+02
## variance..... 5.520265e+03
## var 2:
## best..... 6.450107e+02
## mean..... 6.349600e+02
## variance..... 3.678793e+03
## var 3:
## best..... 7.204414e+02
## mean..... 6.954786e+02
## variance..... 1.132058e+04
## var 4:
## best..... 6.915332e+02
## mean..... 6.668702e+02
## variance..... 9.180323e+03
## var 5:
## best..... 2.366926e+02
## mean..... 2.582784e+02
## variance..... 9.666152e+03
## var 6:
## best..... 1.305135e+02
## mean..... 1.666711e+02
## variance..... 1.335786e+04
## var 7:
## best..... 2.855952e+02
## mean..... 3.080022e+02
## variance..... 1.289387e+04
## var 8:
## best..... 1.677894e+01
## mean..... 4.673058e+01
## variance..... 1.139101e+04
## var 9:
## best..... 1.969262e+02
## mean..... 2.054815e+02
## variance..... 6.896406e+03
##
## GENERATION: 5
## Lexical Fit..... 5.851030e-01 6.314611e-01 6.814275e-01 8.285723e-01 8
.815982e-01 9.206334e-01 9.658675e-01 9.809493e-01 9.999992e-01 9.999992
e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1
.000000e+00 1.000000e+00 1.000000e+00
## #unique..... 106, #Total UniqueCount: 701
## var 1:
## best..... 5.298648e+02
## mean..... 5.269628e+02
## variance..... 6.844983e+03
## var 2:
## best..... 6.450107e+02
## mean..... 6.276543e+02

```

```

## variance..... 7.287974e+03
## var 3:
## best..... 7.204414e+02
## mean..... 7.100831e+02
## variance..... 5.880068e+03
## var 4:
## best..... 6.915332e+02
## mean..... 6.677069e+02
## variance..... 7.487962e+03
## var 5:
## best..... 2.366926e+02
## mean..... 2.623190e+02
## variance..... 1.070744e+04
## var 6:
## best..... 1.305135e+02
## mean..... 1.718447e+02
## variance..... 1.791013e+04
## var 7:
## best..... 2.855952e+02
## mean..... 2.796547e+02
## variance..... 1.341491e+04
## var 8:
## best..... 1.677894e+01
## mean..... 4.121045e+01
## variance..... 1.036078e+04
## var 9:
## best..... 1.969262e+02
## mean..... 1.938870e+02
## variance..... 2.683889e+03
##
## GENERATION: 6
## Lexical Fit..... 5.851030e-01 6.314611e-01 6.814275e-01 8.285723e-01 8
.815982e-01 9.206334e-01 9.658675e-01 9.809493e-01 9.999992e-01 9.999992
e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1
.000000e+00 1.000000e+00 1.000000e+00
## #unique..... 108, #Total UniqueCount: 809
## var 1:
## best..... 5.298648e+02
## mean..... 5.209261e+02
## variance..... 5.523937e+03
## var 2:
## best..... 6.450107e+02
## mean..... 6.419802e+02
## variance..... 2.333251e+03
## var 3:
## best..... 7.204414e+02
## mean..... 7.043907e+02
## variance..... 4.880621e+03
## var 4:
## best..... 6.915332e+02

```

```

## mean..... 6.778627e+02
## variance..... 6.838818e+03
## var 5:
## best..... 2.366926e+02
## mean..... 2.577002e+02
## variance..... 1.015184e+04
## var 6:
## best..... 1.305135e+02
## mean..... 1.748574e+02
## variance..... 1.680571e+04
## var 7:
## best..... 2.855952e+02
## mean..... 2.978389e+02
## variance..... 1.495240e+04
## var 8:
## best..... 1.677894e+01
## mean..... 4.197941e+01
## variance..... 6.674092e+03
## var 9:
## best..... 1.969262e+02
## mean..... 2.137525e+02
## variance..... 9.159039e+03
##
## GENERATION: 7
## Lexical Fit..... 5.851030e-01 6.314611e-01 6.814275e-01 8.285723e-01 8
.815982e-01 9.206334e-01 9.658675e-01 9.809493e-01 9.999992e-01 9.999992
e-01 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1
.000000e+00 1.000000e+00 1.000000e+00
## #unique..... 105, #Total UniqueCount: 914
## var 1:
## best..... 5.298648e+02
## mean..... 5.157806e+02
## variance..... 7.134557e+03
## var 2:
## best..... 6.450107e+02
## mean..... 6.286958e+02
## variance..... 6.967758e+03
## var 3:
## best..... 7.204414e+02
## mean..... 7.083882e+02
## variance..... 6.654024e+03
## var 4:
## best..... 6.915332e+02
## mean..... 6.786248e+02
## variance..... 7.000368e+03
## var 5:
## best..... 2.366926e+02
## mean..... 2.496488e+02
## variance..... 5.335667e+03
## var 6:

```

```

## best..... 1.305135e+02
## mean..... 1.569787e+02
## variance..... 1.222896e+04
## var 7:
## best..... 2.855952e+02
## mean..... 3.069217e+02
## variance..... 1.619635e+04
## var 8:
## best..... 1.677894e+01
## mean..... 4.315991e+01
## variance..... 9.195733e+03
## var 9:
## best..... 1.969262e+02
## mean..... 2.221780e+02
## variance..... 1.082557e+04
##
## 'wait.generations' limit reached.
## No significant improvement in 4 generations.
##
## Solution Lexical Fitness Value:
## 5.851030e-01 6.314611e-01 6.814275e-01 8.285723e-01 8.815982e-01 9.20
6334e-01 9.658675e-01 9.809493e-01 9.999992e-01 9.999992e-01 1.000000e+0
0 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.00
0000e+00 1.000000e+00
##
## Parameters at the Solution:
##
## X[ 1] : 5.298648e+02
## X[ 2] : 6.450107e+02
## X[ 3] : 7.204414e+02
## X[ 4] : 6.915332e+02
## X[ 5] : 2.366926e+02
## X[ 6] : 1.305135e+02
## X[ 7] : 2.855952e+02
## X[ 8] : 1.677894e+01
## X[ 9] : 1.969262e+02
##
## Solution Found Generation 2
## Number of Generations Run 7
##
## Wed Mar 22 22:52:18 2017
## Total run time : 0 hours 3 minutes and 9 seconds

# Now that GenMatch() has found the optimal weights, Let's estimate
# our causal effect of interest using those weights (using Weight.matrix=geno
ut)

mout2 <- Match(Y=Y, Tr=treat, X=X52, estimand="ATT", M=1, replace = TRUE, Wei
ght.matrix=genout2, caliper = 0.45)

```

*#caliper of 0.45 showed better balance, better results!*

`summary(mout2)`

```
##
## Estimate... 1803.6
## AI SE..... 685.2
## T-stat..... 2.6322
## p.val..... 0.0084843
##
## Original number of observations..... 16177
## Original number of treated obs..... 185
## Matched number of observations..... 131
## Matched number of observations (unweighted). 164
##
## Caliper (SDs)..... 0.45 0.45 0.45 0.4
5 0.45 0.45 0.45 0.45 0.45
## Number of obs dropped by 'exact' or 'caliper' 54
```

*#*

*#Let's determine if balance has actually been obtained on the variables of interest*

*#*

```
mb2 <- MatchBalance(treat ~ propensities+age+education+black+ hispanic+ married+
nodegree+ re75+ re74, match.out=mout2, nboots=500)
```

```
##
## ***** (V1) propensities *****
##               Before Matching      After Matching
## mean treatment..... 0.37962      0.35311
## mean control..... 0.0071768      0.35295
## std mean diff..... 135.98      0.058489
##
## mean raw eQQ diff..... 0.36946      0.0048277
## med raw eQQ diff..... 0.35183      0.0021602
## max raw eQQ diff..... 0.83504      0.028075
##
## mean eCDF diff..... 0.49836      0.0072032
## med eCDF diff..... 0.53052      0.0060976
## max eCDF diff..... 0.87124      0.030488
##
## var ratio (Tr/Co)..... 41.251      0.98561
## T-test p-value..... < 2.22e-16      0.8816
## KS Bootstrap p-value.. < 2.22e-16      1
## KS Naive p-value..... < 2.22e-16      1
## KS Statistic..... 0.87124      0.030488
##
##
## ***** (V2) age *****
##               Before Matching      After Matching
## mean treatment..... 25.816      23.885
```

```

## mean control.....      33.225      23.855
## std mean diff.....    -103.55      0.53712
##
## mean raw eQQ diff.....      7.4865      0.40244
## med  raw eQQ diff.....         6         0
## max  raw eQQ diff.....      16         2
##
## mean eCDF diff.....      0.18628      0.015244
## med  eCDF diff.....      0.19692      0.0091463
## max  eCDF diff.....      0.34274      0.054878
##
## var ratio (Tr/Co).....      0.41964      0.90696
## T-test p-value..... < 2.22e-16      0.82857
## KS Bootstrap p-value.. < 2.22e-16      0.846
## KS Naive p-value..... < 2.22e-16      0.96587
## KS Statistic.....      0.34274      0.054878
##
##
## ***** (V3) education *****
##                               Before Matching      After Matching
## mean treatment.....      10.346      10.725
## mean control.....      12.028      10.748
## std mean diff.....     -83.633      -1.4242
##
## mean raw eQQ diff.....      1.7351      0.060976
## med  raw eQQ diff.....         2         0
## max  raw eQQ diff.....         4         1
##
## mean eCDF diff.....      0.090791      0.0060976
## med  eCDF diff.....      0.037581      0.0030488
## max  eCDF diff.....      0.41227      0.030488
##
## var ratio (Tr/Co).....      0.49052      0.94762
## T-test p-value..... < 2.22e-16      0.63146
## KS Bootstrap p-value.. < 2.22e-16      0.96
## KS Naive p-value..... < 2.22e-16      1
## KS Statistic.....      0.41227      0.030488
##
##
## ***** (V4) black *****
##                               Before Matching      After Matching
## mean treatment.....      0.84324      0.82443
## mean control.....      0.073537      0.82443
## std mean diff.....      211.13         0
##
## mean raw eQQ diff.....      0.76757      0
## med  raw eQQ diff.....         1         0
## max  raw eQQ diff.....         1         0
##
## mean eCDF diff.....      0.38485      0

```



```

## med eCDF diff..... 0.38485 0
## max eCDF diff..... 0.76971 0
##
## var ratio (Tr/Co)..... 1.9506 1
## T-test p-value..... < 2.22e-16 1
##
##
## ***** (V5) hispanic *****
## Before Matching After Matching
## mean treatment..... 0.059459 0.045802
## mean control..... 0.072036 0.045802
## std mean diff..... -5.3038 0
##
## mean raw eQQ diff..... 0.016216 0
## med raw eQQ diff..... 0 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.0062883 0
## med eCDF diff..... 0.0062883 0
## max eCDF diff..... 0.012577 0
##
## var ratio (Tr/Co)..... 0.84109 1
## T-test p-value..... 0.47458 1
##
##
## ***** (V6) married *****
## Before Matching After Matching
## mean treatment..... 0.18919 0.18321
## mean control..... 0.71173 0.18321
## std mean diff..... -133.06 0
##
## mean raw eQQ diff..... 0.51892 0
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 1 0
##
## mean eCDF diff..... 0.26127 0
## med eCDF diff..... 0.26127 0
## max eCDF diff..... 0.52254 0
##
## var ratio (Tr/Co)..... 0.75167 1
## T-test p-value..... < 2.22e-16 1
##
##
## ***** (V7) nodegree *****
## Before Matching After Matching
## mean treatment..... 0.70811 0.64885
## mean control..... 0.29584 0.64885
## std mean diff..... 90.437 0
##
## mean raw eQQ diff..... 0.41081 0

```

```

## med   raw eQQ diff.....      0      0
## max   raw eQQ diff.....      1      0
##
## mean  eCDF diff.....      0.20614      0
## med   eCDF diff.....      0.20614      0
## max   eCDF diff.....      0.41227      0
##
## var ratio (Tr/Co).....      0.99753      1
## T-test p-value..... < 2.22e-16      1
##
##
## ***** (V8) re75 *****
##                               Before Matching      After Matching
## mean treatment.....      1532.1      1366.4
## mean control.....      13651      1368.8
## std mean diff.....      -376.45      -0.088536
##
## mean raw eQQ diff.....      12112      163.65
## med   raw eQQ diff.....      13837      0
## max   raw eQQ diff.....      22438      3434.9
##
## mean eCDF diff.....      0.4751      0.016075
## med   eCDF diff.....      0.51248      0.012195
## max   eCDF diff.....      0.6509      0.060976
##
## var ratio (Tr/Co).....      0.12059      0.84094
## T-test p-value..... < 2.22e-16      0.98095
## KS Bootstrap p-value.. < 2.22e-16      0.592
## KS Naive p-value..... < 2.22e-16      0.92063
## KS Statistic.....      0.6509      0.060976
##
##
## ***** (V9) re74 *****
##                               Before Matching      After Matching
## mean treatment.....      2095.6      1828
## mean control.....      14017      1756.3
## std mean diff.....      -243.96      1.8988
##
## mean raw eQQ diff.....      12014      244.86
## med   raw eQQ diff.....      13276      0
## max   raw eQQ diff.....      23256      2719.9
##
## mean eCDF diff.....      0.45911      0.017323
## med   eCDF diff.....      0.50015      0.012195
## max   eCDF diff.....      0.60309      0.079268
##
## var ratio (Tr/Co).....      0.26074      0.90572
## T-test p-value..... < 2.22e-16      0.5851
## KS Bootstrap p-value.. < 2.22e-16      0.314
## KS Naive p-value..... < 2.22e-16      0.68143

```

```
## KS Statistic.....      0.60309          0.079268
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): propensities age education black married nodegree re75 re74
## Number(s): 1 2 3 4 6 7 8 9
##
## After Matching Minimum p.value: 0.314
## Variable Name(s): re74 Number(s): 9
```

## (5.2) Confidence Intervals

```
#estimated treatment effect
est52 = mout2$est ; est52

##           [,1]
## [1,] 1803.562

#confidence intervals
error52 = 1.96*(743) ; error52

## [1] 1456.28

confint52.low <- est52 - error52 ; confint52.up <- est52 + error52 ;
confint52.low ; confint52.up

##           [,1]
## [1,] 347.2819

##           [,1]
## [1,] 3259.842
```

Before Matching Minimum p.value: < 2.22e-16 Variable Name(s): propensities age education black married nodegree re75 re74 Number(s): 1 2 3 4 6 7 8 9

After Matching Minimum p.value: 0.532 Variable Name(s): re74 Number(s): 9

---

Estimate... 1782.4 AI SE..... 689.6 T-stat..... 2.5848 p.val..... 0.0097449

Original number of observations..... 16177 Original number of treated obs..... 185  
 Matched number of observations..... 131 Matched number of observations  
 (unweighted). 164

Caliper (SDs)..... 0.45 0.45 0.45 0.45 0.45 0.45 0.45 0.45 0.45 Number of  
 obs dropped by 'exact' or 'caliper' 54

*Finally, Success!* Our Multivariate Genetic Matching **indeed** increase our lowest p-values, increased many of the other p values, and reduced the number of "poorly balanced" variables from 7 to 1(!).

This makes sense, since in the propensity scores matching we didn't utilize the power of Genetic Matching, but now that we do, we see how it improves our balance much more than in previous attempts: normal Match() function using propensity scores or multivariate matching, or GenMatch() for propensity scores.

This increases the balance we have, thus makes the results more credible, and our final model for estimation less sensitive.

After trying various calipers and finding an optimal point where less units are dropped but there is still balance, Caliper of 1 seemed well balanced. We get the results of: Estimate... 1782.8

Estimate... 1782.8 AI SE..... 942.79 T-stat..... 1.891 p.val..... 0.058626

Original number of observations..... 16177 Original number of treated obs..... 185  
Matched number of observations..... 177 Matched number of observations  
(unweighted). 210

Caliper (SDs)..... 1 1 1 1 1 1 1 1 Number of obs dropped by 'exact' or  
'caliper' 8

## 6: Experimenting with the genmatch and match functions

(By modifying and adding arguments, learning how they work, and how they impact analysis. Key arguments include ???M???, ???pop size???, ???estimand???, ???exact???, ???caliper???, ???replace???, etc).

### Caliper

I've experimented with the caliper in Match and GenMatch functions, as talked about more in length in question 4. Starting from Caliper=0.0 (exact matching), I've slowly, iteratively, increased the caliper to find a balanced point where not too many observations are dropped. What happens is that the caliper (or "maximum difference distance between matched units" allowed), if set to 0 or some very small amount, will find only exactly or very closely matched units, and drop the rest. Therefore, it might change the underlying structure of the treated group taken into consideration - for example, it may only find good matches to Black, 30 year-old people with No Degree. That would decrease our external validity and generalizability, since our results would only apply to that subpopulation. In the cases I checked here, it seemed that usually, when I set the caliper to 0.0, the resulting ATT was further away from the experimental results (ATE). As the caliper increased approaching the point of dropping only ~10% of units or less, I saw that the Matched results approach the experimental results. This happens since if we limit the compared subpopulations too much, we miss information. We don't want to eliminate much of our precious "treatment" units information, and want to consider a wider set of treated sample so we'd have more external validity and generalizability. However, for genetic matching, increasing the caliper incrementally until even over 1.3, the performance of the balance and matching was worse than without specifying caliper. When Genetic Matching doesn't use caliper, it uses distance.tolerance, which performed better.

## Replace

Replace sets whether matching should be done with replacement or not. When I set `replace=False`, that reduced the quality of the results, resulting in poorer resulting matching - most p-values were not increased by as much. This happens since if we remove observations from the pool every time we match them, we decrease the pool size from which we can choose the best match for the next units. Therefore, that same observation which was dropped might have been the best match for next 3 units, but we will have to suffice with increasingly worse matches each time. So usually Replace decreases the quality of matching, and this was visible in this example by not increasing as much many of the P-Values.

## Pop Size, Generations

The Population Size and Generations each determine how deep / wide should our genetic search be. Therefore, increasing these improves our final result, but takes considerably more time. As I increased the Population size and generations of GenMatch, it took much longer, but overall normally improved the results - seeing with better improved p-values, and sometimes even a smaller AI Standard Error (although I'm not sure it relates specifically to this).

## Estimand

Estimand: we are interested in estimating the ATT - the sample average treatment effect for the treated. Otherwise the estimand could be: "ATE" - the sample average treatment effect, or "ATC" - the sample average treatment effect for the controls.

Using ATE *completely* changed the results, to a NEGATIVE, but insignificant effect. Estimate... -509.14 AI SE..... 495.11 The confidence interval crosses 0 to the positive side. But this is not the measure of our interest, since we want the treatment effect FOR THE TREATED; therefore we want to take the treated units and match controls onto them, measuring the difference between these pairs. But ATE considers the entire control group, instead of focusing on the treated, therefore we have differences which aren't most representative of the actual effect of our treatment.

We also didn't improve our balance by much this way. We remained with 3 variables with the unimproved lowest p-value result.

Before Matching Minimum p.value: < 2.22e-16 Variable Name(s): propensities age education black married nodegree re75 re74 Number(s): 1 2 3 4 6 7 8 9

After Matching Minimum p.value: < 2.22e-16 Variable Name(s): propensities age education re75 re74 Number(s): 1 2 3 8 9

## M

M=1 is one-to-one matching, which is also the default. Increasing this will make matching into many-to-one matching. In my experiments, increasing the M DECREASED the quality of covariance balance after Matching. Increasing the M up to 5, decreased the improvement in

p-value after matching up to the point of having no improvement in the lowest p-value. Moreover, the resulting estimate was *significantly far away* from the experimental results and even from any other method: Estimate... 1014.2 AI SE..... 620.31 .