Algorithm to estimate "simple matching estimators" as presented in Abadie, Drukker, Leber Herr & Imbens (2004): Implementing matching estimators for average treatment effects in Stata. The Stata journal, 4(3), 290-311

When trying to find matches for the observations in the treatment group using covariates, follow these steps:

1.- calculate the absolute distance between the covariates for each observation in the treatment group and all the observations in the control control group
2.- those that are the closest, are your match, and are selected
3.- if more than one match, calculate the average of the outcomes for the matched observations
4.- subtract the outcome for the treatment observation being matched from the average of the outcomes estimated in step 3, or the outcome of the closest match found in step 2. Repeat for all the treatment observations

Once you have found matches for all the observations in the treatment group:
5.- calculate the average of the differences

Note: we follow the same steps to find the matches for the control group

Example:

obs_i	$group_i$	cov_i	y_i
1	0	2	7
2	0	4	8
3	0	5	6
4	1	3	9
5	1	2	8
6	1	3	6
7	1	1	5

Where group = 0 control, 1 = treatment
cov = covariate (to be used for matching)
y = outcome

Next, we will select the matches for the treatment (group = 1) group:

1.- calculate the absolute distance between the covariates for each observation in the treatment group and all the observations in the control group
and
2.- those that are the closest, are your match, and are selected

for observation 4 (cov = 3), select observations 1 & 2, because their distances are the closest (1)

$diff1 := |3 - 2|$ $diff1 = 1$
$diff2 := |3 - 4|$ $diff2 = 1$
$diff3 := |3 - 5|$ $diff3 = 2$

$diff4 := |2 - 2|$ $diff4 = 0$
$diff5 := |2 - 4|$ $diff5 = 2$
$diff6 := |2 - 5|$ $diff6 = 3$

for observation 5 (cov = 2), select control observation 1, because its distance is the closest (0)

for observation 6 (cov = 3), select observations 1-2, because their distance is the closest (1)

$diff7 := |3 - 2| \quad diff7 = 1$

$diff8 := |3 - 4| \quad diff8 = 1$

$diff9 := |3 - 5| \quad diff9 = 2$

$diff10 := |1 - 2| \quad diff10 = 1$

$diff11 := |1 - 4| \quad diff11 = 3$

$diff12 := |1 - 5| \quad diff12 = 4$

for observation 7 (cov = 1), select observation 1, because its distance is the closest (1)

3.- if more than one match, calculate the average of the outcomes for the matched observations

observation 4: $match4 := \frac{7 + 8}{2} \quad match4 = 7.5$

observation 6: $match6 := \frac{7 + 8}{2} \quad match6 = 7.5$

for observations 5 and 7, we do not need to calculate this step

4.- subtract the outcome for the treatment observation being matched from the average of the outcomes estimated in step 3, or the outcome of the closest match found in step 2. Repeat for all the treatment observations

$outc\_d1 := 9 - 7.5 \quad outc\_d1 = 1.5$

$outc\_d2 := 8 - 7 \quad outc\_d2 = 1$

$outc\_d3 := 6 - 7.5 \quad outc\_d3 = -1.5$

$outc\_d4 := 5 - 7 \quad outc\_d4 = -2$

5.- calculate the average of the differences

$$\frac{outc\_d1 + outc\_d2 + outc\_d3 + outc\_d4}{4} = -0.25$$

Because we are only calculating the average for the individuals in the treatement group, this effect type is what is known as the **A**verage **T**reatment for the **T**retated (ATT)

We repeat the procedure for the observations in the control (group=0) group:

1.- calculate the absolute distance between the covariates for each observation in the control group and all the observations in the treatment group
and
2.- those that are the closest, are your match, and are selected

for observation 1 (cov = 2), select observation 5, because its distance is the closest (0)

$diff1 := |2 - 3| \quad diff1 = 1$

$diff2 := |2 - 2| \quad diff2 = 0$

$diff3 := |2 - 3| \quad diff3 = 1$

$diff4 := |2 - 1| \quad diff4 = 1$

$diff5 := |4 - 3| \quad diff5 = 1$

$diff6 := |4 - 2| \quad diff6 = 2$

$diff7 := |4 - 3| \quad diff7 = 1$

$diff8 := |4 - 1| \quad diff8 = 3$

for observation 2 (cov = 4), select observation 4 and 6, because their distance is the closest (1)

for observation 3 (cov = 5), select observations 4 and 6, because their distance is the closest (2)

$diff9 := |5 - 3|$ $diff9 = 2$

$diff10 := |5 - 2|$ $diff10 = 3$

$diff11 := |5 - 3|$ $diff11 = 2$

$diff12 := |5 - 1|$ $diff12 = 4$

3.- if more than one match, calculate the average of the outcomes for the matched observations

observation 1: $match1 := \frac{9 + 6}{2}$ $match1 = 7.5$

observation 3: $match3 := \frac{9 + 6}{2}$ $match3 = 7.5$

for observation1, we do not need to calculate this step

4.- subtract the outcome for the TREATMENT observation being matched from the average of the outcomes estimated in step 3, or the outcome of the closest match found in step 2. Repeat for all the control observations

$outc\_d5 := 8 - 7$ $outc\_d5 = 1$

$outc\_d6 := 7.5 - 8$ $outc\_d6 = -0.5$

$outc\_d7 := 7.5 - 6$ $outc\_d7 = 1.5$

NOTICE that we always subtract the outcomes from the TREATMENT GROUP from those of the CONTROL GROUP. Even when we are doing this for the CONTROL group

5.- calculate the average of the differences

$$\frac{outc\_d5 + outc\_d6 + outc\_d7}{3} = 0.66667$$

Because we are only calculating the average for the individuals in the control group, this effect type is what is known as the **A**verage **T**reatment for the **C**ontrol (ATC)

Finally, we calculate the Average the Treatment Effect:

$$\frac{outc\_d1 + outc\_d2 + outc\_d3 + outc\_d4 + outc\_d5 + outc\_d6 + outc\_d7}{7} = 0.14286$$

Example in handout but using program "Matching" in R

```
> setwd("~/propensity scores/book/Chapter6")
> library("foreign", lib.loc="C:/Program Files/R/R-2.13.2/library")
> library("Matching", lib.loc="C:/Program Files/R/R-2.13.2/library")
Loading required package: rgenoud
##  rgenoud (Version 5.7-3, Build Date: 2011-05-04)
##  See http://sekhon.berkeley.edu/rgenoud for additional documentation.
Loading required package: MASS
##
##  Matching (Version 4.8-0, Build Date: 2012/01/19)
##  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 for R.''
##   Journal of Statistical Software, 42(7): 1-52.
##

> data <- read.dta("ExampleSimpleMatchingEstimator.dta")
> data
  w x y
1 0 2 7
2 0 4 8
3 0 5 6
4 1 3 9
5 1 2 8
6 1 3 6
7 1 1 5

> S_ATE <- Match(Y=data$y, Tr=data$w, X=data$x, estimand="ATE")
> summary(S_ATE)

Estimate...  0.14286
AI SE......  0.91382
T-stat.....  0.15633
p.val......  0.87577

Original number of observations..............  7
Original number of treated obs...............  4
Matched number of observations...............  7
Matched number of observations  (unweighted).  11

> S_ATT <- Match(Y=data$y, Tr=data$w, X=data$x, estimand="ATT")
> summary(S_ATT)

Estimate...  -0.25
AI SE......  1.0542
T-stat.....  -0.23715
p.val......  0.81254

Original number of observations..............  7
Original number of treated obs...............  4
Matched number of observations...............  4
Matched number of observations  (unweighted).  6

> S_ATC <- Match(Y=data$y, Tr=data$w, X=data$x, estimand="ATC")
> summary(S_ATC)

Estimate...  0.66667
AI SE......  0.60349
T-stat.....  1.1047
p.val......  0.26929

Original number of observations..............  7
Original number of control obs...............  3
Matched number of observations...............  3
Matched number of observations  (unweighted).  5
```

Average Treatment Effect (ATE)

Average Treatment for the Treated (ATT)

Average Treatment for the Control (ATC)