



# Introduction to Optimalpostmatch package in R

**Priyalatha Govindasamy**  
(gpriyalatha@hotmail.com)

**Antonio Olmos**  
(antonio.olmos@du.edu)

Research Methods & Statistics  
University of Denver



# Paper overview

- Brief introduction to Propensity score matching
- Types of Propensity score matching
- Overview of Optimal matching
- Outcome analysis after Optimal matching
- Introduction to Optimalpostmatch package in R

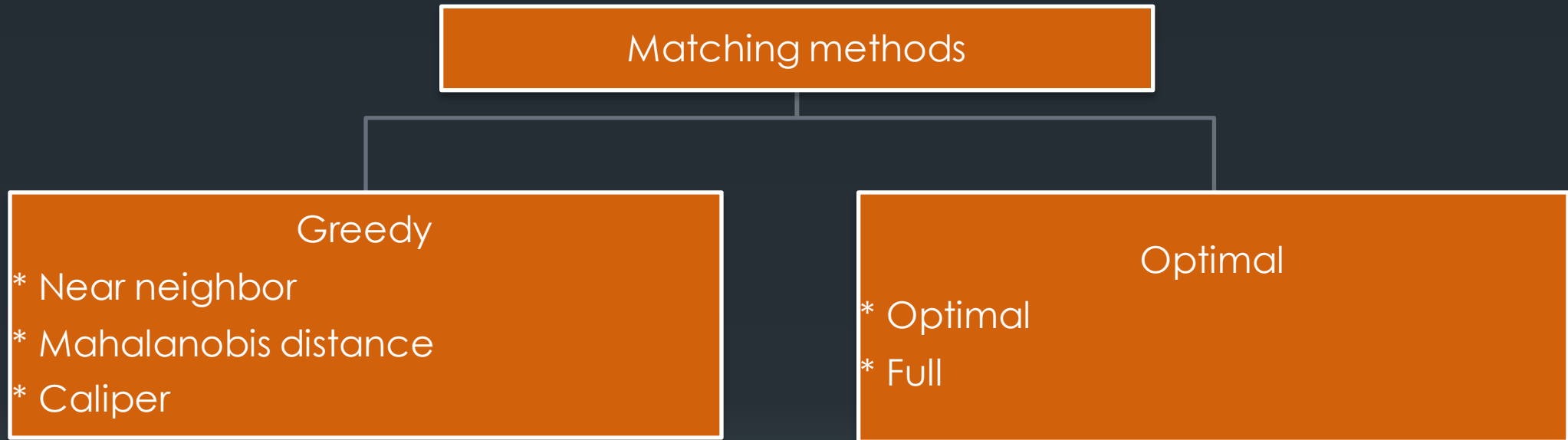
# What is Propensity score matching?

- Creates matched treatment and control groups that are as similar as possible based on a wide range of observed covariates (Stone & Tang, 2013)
- Upon computing and matching the groups on Propensity scores, the only differences between the treatment and control group should be the reflection of whether the groups have received treatment or not.
- Researchers could conclude that any observed differences in the outcome is the result of the treatment.

# How to perform Propensity Score Matching?

- Different matching methods are used to create matched pairs of treated and non-treated participants/observations
- Generally the matching methods can be grouped into (1) greedy and (2) optimal

# How to perform Propensity Score Matching?



## **Greedy versus Optimal :**

Literature states that optimal matching methods provides better matched group compared to greedy methods (Guo & Fraser, 2015)

# What is Optimal matching?

- In optimal matching:

Step 1: It uses network flow theory to optimize the selection of individuals by minimizing the distances between the matched pairs.

Step 2: The overall difference between the treated and non-treated groups from the matched data will be computed.

Step 3: The matched pairs will be altered until the total minimum distance is achieved.

# Why Optimal matching is better?

- In optimal matching, decision about selecting pairs are dependent on the earlier matches.
- Selection of matching pairs are not only related to the smallest distance between the pairs but also dependent on the matching of other pairs.
- Optimizing the differences between the overall matches to be the smallest is the advantage of optimal over greedy matching.

# Matched data from Optimal matching method

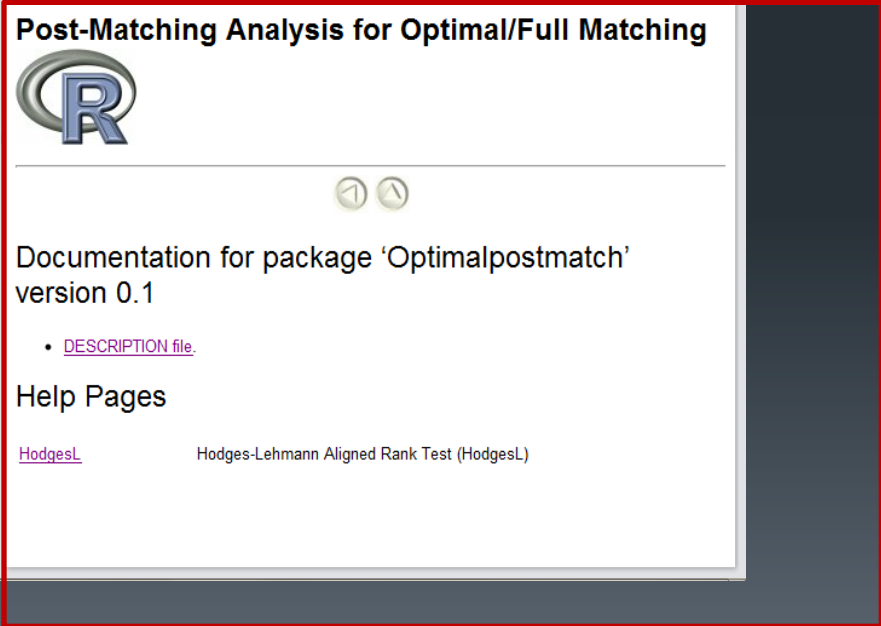
- Optimal matching algorithm develops strata containing matched observations for treated and non-treated cases that are as homogenous as possible.
- The creation of strata makes the observations dependent of one another.
- Also, the selection of pairs are dependent on other pairs in the data.
- The creation of matched pairs using optimal approach creates dependency between the observations/cases.
- Dependency in the observations violates the independence assumption that is crucial for regular multivariate analyses.



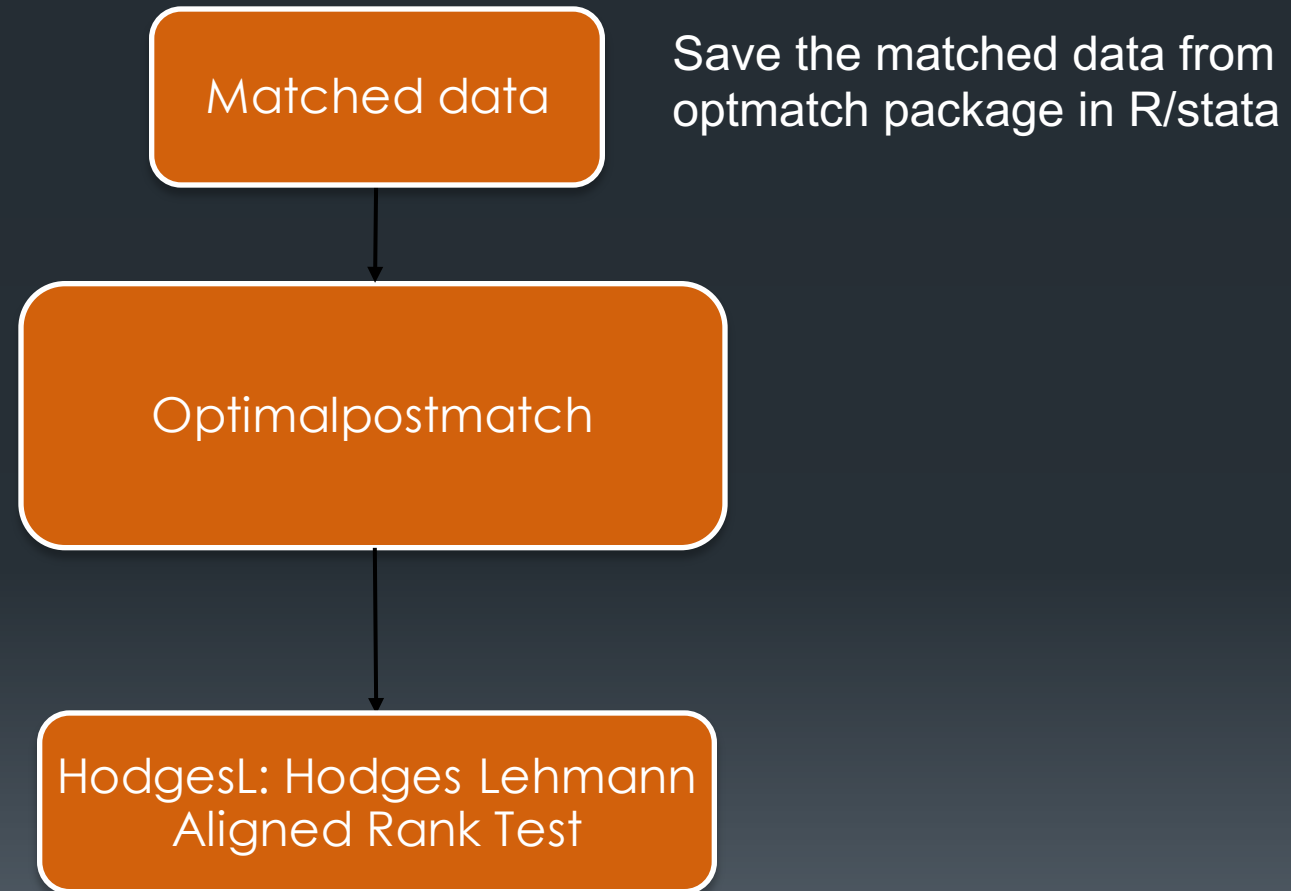
# Option: Hodges-Lehmann Aligned Rank test



- Hodges Lehman Aligned Rank Test is a non-parametric test for ranks
- It estimates the Average Treatment Effect (ATE) by assessing the weighted average of the mean differences between treated and control participants of all matched sets within the sample.
- Then, Hodges-Lehmann aligned rank test is used as a significance test for the computed ATE (Guo & Fraser, 2015).
- There is no current implementation of the Hodges-Lehmann Aligned Rank test in R



# Demonstration of how to use the package



# Demonstration of how to use the package (cont..)

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for setting the working directory, loading data, and installing the `Optimalpostmatch` package. Lines 6-15 are highlighted with a red box.
- Environment:** Shows the `match.data` object with 14 observations and 4 variables.
- Files:** A file explorer showing the project directory structure, including `example.csv` and `Example_using_Optimalpostmatch.Rmd`. A red box highlights the `match.data` object in the Environment pane and the `example.csv` file in the Files pane.
- Console:** Displays the output of the R code, including the installation of the `Optimalpostmatch` package and the results of the `match.data` object.

```
1 ---
2 title: "Example of Optimalpostmatch"
3 author: "Priya"
4 date: "Thursday, March 26, 2015"
5 output: word document
6 ---
7 ## Setting the working directory and loading the datafile
8 {r}
9 setwd("G:/R_package_Optimalpost/Document for R package/Documents for testing the package")
10
11 match.data <- read.csv("example.csv")
12
13 # Required library
14 library("Optimalpostmatch")
15
16 ## Difference in Difference test
17
30:1 Chunk 3
```

Console output:

```
** testing if installed package can be loaded
*** arch - i386
*** arch - x64
* DONE (Optimalpostmatch)
> setwd("G:/R_package_Optimalpost/Document for R package/Documents for testing the package")
> setwd("G:/R_package_Optimalpost/Document for R package/Documents for testing the package")
> match.data <- read.csv("example.csv")
> # Required library
> library("Optimalpostmatch")
> names(match.data)[names(match.data)=="score09"]="y" #Rename the outcome variable
> names(match.data)[names(match.data)=="treatm"]="treat" #Rename the treatment variable
> Hodgesl(match.data, 14)
$ATE
[1] 9.75

$`z scores`
[1] 1.798119

$p-value`
[1] "nonsignificant"

> |
```

**Step 1:** Load the matched data into R

**Step 2:** Load the **Optimalpostmatch** package into R

# Demonstration of how to use the package (cont..)

## Conducting Hodges-Lehmann Aligned Rank Test

```
23 ^
24 ▾ ## Hodges Lehman Aligned Rank Test
25 ▾ ```{r}
26 names(match.data)[names(match.data)=="score09"]="y" #Rename the outcome variable
27 names(match.data)[names(match.data)=="treatm"]="treat" #Rename the treatment variable
28
29 HodgesL(match.data, 14)
30 ^```
```

### Step 3:

- i. Rename the outcome variable
- ii. Rename the treatment variable
- iii. Run the analysis

# Demonstration of how to use the package (cont..)

## Hodges-Lehmann Aligned Rank Test Output

Console G:/R\_package\_Optimalpost/Document for R package/Documents for testing the package/ ↗

```
> ## Hodges Lehman Aligned Rank Test  
> names(match.data)[names(match.data)=="score09"]="y"      #Rename the outcome variable  
> names(match.data)[names(match.data)=="treatm"]="treat"   #Rename the treatment variable  
> HodgesL(match.data, 14)
```

```
$ATE  
[1] 9.75  
  
$`z scores`  
[1] 1.798119  
  
$`p-value`  
[1] "nonsignificant"
```

**ATE = Average  
Treatment Effect**

Z scores for the  
computed ATE

Significance test on  
the computed ATE

# Limitations

- The package is still under development
- \* Exploring more statistical analysis that are suitable for matched data created using optimal matching
- \* Difficult to find literature on the application of optimal matching

# Conclusion

- Package provides practitioners with an easy-to-use alternative to compute the Hodges-Lehmann Aligned Rank test
- This package connects with the `optmatch` by Hansen et al., 2015
- These developments will hopefully increase the use of optimal matching





# Where to find the package?

- The package can be downloaded from:

<https://portfolio.du.edu/Antonio.Olmos/page/54952>

Thank you