

# Lab 7: Matching

## *Advanced Applied Statistics*

### Description

In this lab session we will focus on preprocessing our data using different matching techniques. In the previous lab sessions we have included all observations at our disposal to infer causality. In this lab session we will focus on matching observations and thus limit our analysis to observations for whom we have reasonable matches.

### I. Install the packages

We are going to use a variety of packages. The packages listed in `matching.R` (from line 5 to 18) are required to run the remainder of the file (or will be loaded automatically by the required packages). Use the `install.packages()` function to install the packages.

### II. Inspect the `lalonde` data

The data used in this lab session (and in multiple R packages) is originally used in LaLonde (1986) and analyzed with propensity score matching in Dehejia and Wahba (1999).

Variable	Description
<code>treat</code>	an indicator variable for treatment status.
<code>age</code>	age in years.
<code>educ</code>	years of schooling.
<code>black</code>	indicator variable for blacks.
<code>hisp</code>	indicator variable for Hispanics.
<code>married</code>	indicator variable for marital status.
<code>nodegr</code>	indicator variable for high school diploma.
<code>re74</code>	real earnings in 1974.
<code>re75</code>	real earnings in 1975.
<code>re78</code>	real earnings in 1978. ( <i>dependent variable</i> )
<code>u74</code>	indicator variable for earnings in 1974 being zero.
<code>u75</code>	indicator variable for earnings in 1975 being zero.

You are going to load the data set and inspect it before further use.

1. Get an overview of the different variables in the `lalonde` data frame.
2. Get summary statistics for all the variables.
3. Copy the data frame to a new object (data frame) called `lal`.

### III. Calculate and compare propensity scores

1. Estimate a logistic regression with `treat` as the dependent variable.
2. Predict and save the propensity score for each unit.

3. Use the `matching()` function from the `arm` package to process the data with the propensity scores.
4. Compare the overlap between the treatment and control group before and after matching.
5. Compare the balance between the treatment and control group before and after matching.

#### IV. Average effects for unmatched and matched units

1. Estimate the average treatment effect of `treat` on `re78` using the unmatched and matched samples.

#### V. Average effects using different matching methods

1. Use the `MatchIt` package to match units with nearest neighbor matching, genetic matching, optimal matching and coarsened exact matching.
2. Estimate the average treatment effect on the treated with the different matches in simple OLS regressions (hint: the weights from the matching procedures weigh the control group to look like the treated group and, in other words, enables us to estimate ATET).
3. Include covariates as well in the OLS regressions (*not provided in the script*).
4. Create a data frame with the effect estimates and illustrate the effects from the different matching methods in a figure.

#### VI. Export the data set and import it in STATA

1. Use `write.csv()` to export the `lal` data frame.
2. Import the data file to STATA.
3. Use STATA to estimate ATE and ATT of `treat` on `re78` using nearest neighbor matching. (hint: `teffects nnmatch` or `psmatch2`)
4. How are the estimates compared to the results from R?