

stat_match

Paolo Chirico

Version 1.0

“stat_match” is a package to perform statistical matching algorithms by gretl. The actual version of the package allows to combine two matching methods with two proximity measures (Table 1).

Table 1: Matching procedures by “stat_match”

Statistical Matching procedures		matching method	
		radius matching	one-to-one matching
proximity measure	propensity score	radius matching with propensity score	one2one matching with propensity score
	mahalanobis distance	radius matching with mahalanobis distance	one2one matching with mahalanobis distance

In the following sections:

- Section 1. reports some brief methodological notions on statistical matching;
- Section 2. Illustrates the functions of the package;
- Section 3. shows a simple example.

1. Brief methodological notions on statistical matching

One goal of causal inference is to estimate the *Average Treatment effect on Treated (ATT)*:

$$ATT = E \left[Y_i^{(1)} | T_i = 1 \right] - E \left[Y_i^{(0)} | T_i = 1 \right] \quad (1)$$

where $Y_i^{(1)}$ represents an outcome (e.g. an exam score) observable on a generic unit, i , of the population if it received a treatment (e.g. additional lessons); $Y_i^{(0)}$ represents the outcome in case the unit does not receive the treatment; T_i identifies whether unit i actually receives the treatment ($T_i = 1$) or not ($T_i = 0$).

The main problem in estimating ATT is that $Y_i^{(0)}$ is not observable on the treated units, therefore $E \left[Y_i^{(0)} | T_i = 1 \right]$ must be estimated indirectly by making appropriate assumptions.

If we assume the existence of a set of covariates X , having the following property in the population:

$$E \left[Y_i^{(0)} | T_i = 1, X_i = x \right] = E \left[Y_i^{(0)} | T_i = 0, X_i = x \right] \quad \forall x \quad (2)$$

then units with equal X value have equal $E \left[Y_i^{(0)} | X_i = x \right]$ independently of the actual treatment status.

Under the condition (2), ATT could be estimated matching each treated unit, i , of a sample with one or more untreated units showing the same X value. More specifically, named $y_i^{(0)}$ the Y value of the untreated unit matched with i ¹, ATT can be estimated as:

$$\widehat{ATT} = \bar{y}_{S_1} - \bar{y}_{S_1}^{(0)} \quad (3)$$

where S_1 is the set of the treated units².

Nevertheless, it is very rare to observe treated and untreated units having exactly the same values of X in observational data samples!

To apply the estimator (3) it is necessary to match units showing X values which are close although not coincident. That is the aim of the statistical matching procedures: matching units which are similar with respect to a set of covariates. The proximity is generally measured using two alternative types of proximity measures:

- absolute difference in propensity score
- mahalanobis distance

The propensity score corresponds to the probability:

$$\phi_i = P(T_i = 1|x_i) \quad (4)$$

which is usually estimated with a probit or logit model; the mahalanobis between i and i' is:

$$d_M(i, i') = (x_i - x_{i'})'S(x_i - x_{i'}) \quad (5)$$

where x_i and $x_{i'}$ are the vectors of the observations of X on i and i' ; S is the sample covariance matrix of the covariates.

Since the matching involves units having close but not coincident values of X , the matching must be validated with tests (balance tests) that verify the null hypothesis: $f(X_i|T_i = 1) = f(X_i|T_i = 0)$.

2. The `stat_match` package

At present (version 1.0), `stat_match` allows to perform two common matching methods:

- radius matching: each treated unit is matched with untreated units whose proximity measure is less than a predetermined limit (caliper);
- one-to-one matching: each treated unit is matched with the nearest untreated unit, in terms of proximity measure.

These methods (available with both propensity score and mahalanobis distance) are performed by calling the `stat_match()` function, as follows:

```
bundle BN = stat_match(T, Y, X, method, measure, caliper, common, verbose)
```

where:

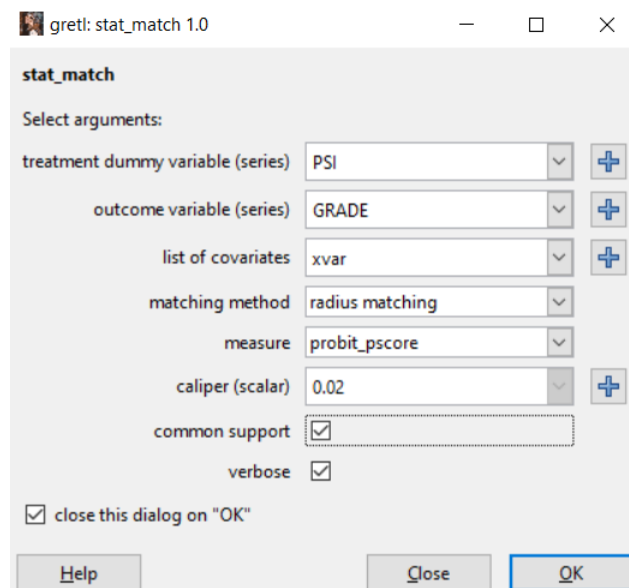
- `BN` is the name of the bundle that will contain the results (see Section 2.3)
- `T` (series) is the treatment dummy variable

¹ If two or more untreated units are matched with i , $y_i^{(0)}$ is the outcome sample mean of these units.

² It is clear that: $\bar{y}_{S_1} = \widehat{E}[Y_i^{(1)}|T_i = 1]$ and $\bar{y}_{S_1}^{(0)} = \widehat{E}[Y_i^{(0)}|T_i = 1]$.

- `Y` (series) is the outcome variable
- `X` (list) is the list of covariates
- `method` (int) is an integer, from 1 to 2, to set the matching method:
 1. radius matching;
 2. one-to-one matching;
- `measure` (int) is an integer, from 1 to 3, to set the type of proximity measure:
 - 1 = propensity score by probit regression;
 - 2 = propensity score by logit regression;
 - 3 = mahalanobis distance;
- `caliper` (scalar) is the maximum proximity measure to make the matching; it must be set if radius method is used, but it can be omitted in case of one-to-one method (see section 2.1 for more details)
- `common` (bool) is a boolean value (0/1) to restrict matching to only treated units that have propensity scores in the same range as non-treated units (`common=1`); it is not used in case of mahalanobis distance.
- `verbose` (bool) is a boolean value (0/1) to set either full visual output or minimal visual output.

Using GUI, the following window will open:



2.1 The caliper setting

If the caliper is not set, it assumes the default value (zero), which is not a realistic value, but just a trick for:

- a. "not using any caliper" in case of one-to-one method;
- b. "having caliper set automatically" in case of radius method: $\text{caliper} = 0.2 \times \text{standard deviation of propensity scores}$ if propensity score is set (Austin, 2011); $\text{caliper} = \text{mean mahalanobis distance from centroid}$ if mahalanobis distance is set.

These values are only a starting point for setting up the most appropriate matching procedure. It is advisable to repeat the procedure with values lower and higher than those in order to identify the matching which guarantees the best balance of the covariates.

2.2 Bundle objects

The `stat_match` function creates a bundle that contains the following objects:

scalars:

ATT = estimate Average Treatment on Treated
FATT = False ATT = $m_1 - m_0$
 m_0 = mean of Y in non-treated units
 m_{0_match} = mean of Y in non-treated and matched units
 m_1 = mean of Y in treated units
 m_{1_match} = mean of Y in treated and matched units
 n_0 = number of non-treated units
 n_{0_supp} = number of non-treated units on the (common) support
 n_{0_match} = number of non-treated and matched units
 n_1 = number of treated units
 n_{1_supp} = number of treated units on the (common) support
 n_{1_match} = number of treated and matched units
radius = value of the caliper radius
 s_{20} = variance of Y in non-treated units
 s_{20_match} = variance of Y in non-treated and matched units
 s_{21} = variance of Y in treated units
 s_{21_match} = variance of Y in treated and matched units
 s_2 = pooled variance of Y
SE_ATT = standard error of ATT
SE_FATT = standard error of FATT

series:

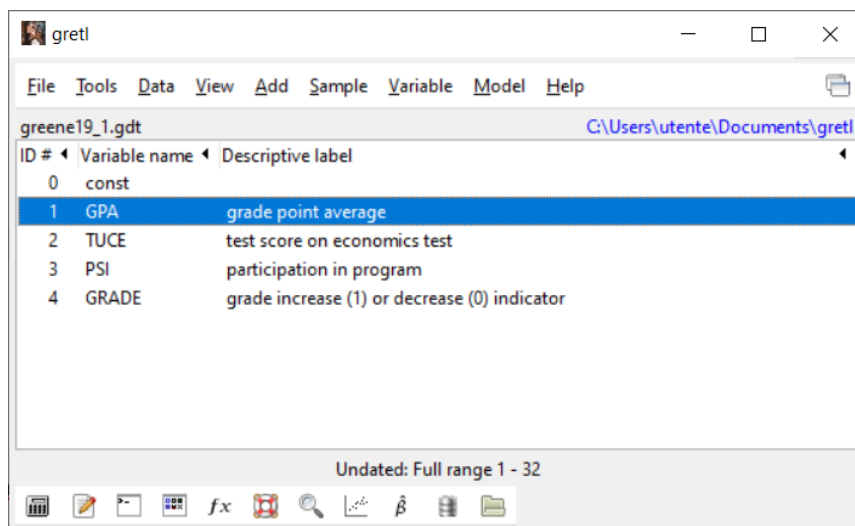
ctrl = counterfactual $y_i^{(0)}$ for treated units
nt1 = number of matched non-treated units for each treated and matched units
outcome = outcome variable (Y)
pscore = propensity score (optional)
support = dummy to identify the matched units
treatment = treatment variable (T)
match = id of the untreated unit, matched with the i treated unit (optional)
weight = weights used to compute the sample means in Formula (3)

other object:

test = matrix of some results of balance test
type_support = string reporting the type of support
X = list of covariates

3. Example

In this section we report a simple example with the dataset `greene19_1.gdt`:



We estimate the *ATT* of a program for teaching economics on the improvements of students' grades. In this case, the outcome variable is GRADE, the treatment variable is PSI, and GPA and TUCE are the covariates. A radius matching is performed with probit propensity score, radius=0.02, and common support.

The script is:

```
open greene19_1.gdt
include stat_match.gfn
list xvar = GPA TUCE
bundle BN = stat_match(PSI, GRADE, xvar, 1, 1, 0.02, 1, 1)
```

If `verbose=1` is set, the `stat_match` function produces the following output:

PROBENSITY SCORE by probit regression:
Dependent variable (trmmt): PSI

Probit, using observations 1-32
Dependent variable: trmmt
Standard errors based on Hessian

	coefficient	std. error	z	p-value
const	-0.956684	1.70862	-0.5599	0.5755
GPA	-0.0117638	0.520928	-0.02258	0.9820
TUCE	0.0380284	0.0635424	0.5985	0.5495

Mean dependent var 0.437500 S.D. dependent var 0.504016
McFadden R-squared 0.009380 Adjusted R-squared -0.127419
Log-likelihood -21.72435 Akaike criterion 49.44870
Schwarz criterion 53.84591 Hannan-Quinn 50.90625

Number of cases 'correctly predicted' = 19 (59.4%)
f(beta*x) at mean of independent vars = 0.394
Likelihood ratio test: Chi-square(2) = 0.411407 [0.8141]

		Predicted	
		0	1
Actual	0	17	1
	1	12	2

Test for normality of residual -
Null hypothesis: error is normally distributed
Test statistic: Chi-square(2) = 0.0463805
with p-value = 0.977077

STATISTICAL MATCHING:
method: radius method
type measure: Propensity Score
caliper = 0.020
type support: common support
outcome variable : GRADE
treatment variable: PSI

SAMPLE MEANS:

units	Treated	Untreated	Difference	SE(Diff.)	t-radio
unmatched	0.5714	0.1667	0.4048	0.1584	2.56
matched	0.6154	0.1679	0.4474	0.1745	2.56

Average Treatment effect on the Treated: ATT = 0.4474 (0.1745)

units	Untreated	Treated
Off support	0	1
On support	18	13
-unmatched	1	0
-matched	17	13
Total	18	14

BALANCE TESTS:

Covariate	M(T)	M(C)	%bias	t-test	V(T)/V(C)
GPA	3.157	3.185	-5.81	-0.16	1.72
TUCE	23.077	23.072	0.13	0.00	0.97

Significance of the tests: 5%

Equality means test: * if $abs(t) > 1.96$

Equality variances test: * if variance ratio is out of [0.31, 3.28]

The output above consists of three parts:

1. Propensity score estimation with a logit/probit regression (this part does not appear in case mahalanobis distance is used)
2. Statistical Matching
3. Matching validation by balance tests.

The evaluation of probit and logit regressions for the propensity score is a controversial issue, which we do not address in this guide. Validation of the whole procedure is deferred to part 3.

Part 2. reports the matching results: we can note that there is a difference between the causal effect estimate using matching (44.7%) and not using matching (40.8%); if matching is validated, the corresponding estimate (44.7%) should be considered less biased than one without matching (40.8%).

The balance test (part 3. of the output) tells us that the difference between the means of the covariates in the group of matched treated units (T) and matched non-treated units (control group, C) is not significant. Same result for the covariate variances. The estimate of the standardized percentage bias (Rosenbaum and Rubin, 1985) is also reported; the balance is generally acceptable if the standardized percentage bias is less than 10%. On the basis of these results the matching should be validated.

References

1. Austin, P.C. "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies." *Pharmaceutical statistics* 10.2: 150-161 (2011).
2. Heckman, J.J., Ichimura, H., Todd, P.: Matching as an econometric evaluation estimator. *The review of economic studies*, 65(2), 261-294 (1998)
3. Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41-55 (1983)
4. Rosenbaum, Paul R., and Donald B. Rubin. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *The American Statistician* 39.1 (1985): 33-38.

5. Rubin, D.B.: Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* 66(5):688 (1974)
6. Rubin, D. B.: Bias reduction using Mahalanobis-metric matching. *Biometrics*, 293-298 (1980)