

tebalance box — Covariate balance box

[Description](#)
[Options](#)[Quick start](#)
[Remarks and examples](#)[Menu](#)
[Reference](#)[Syntax](#)
[Also see](#)

Description

`tebalance box` produces box plots that are used to check for balance in matched samples after `teffects nnmatch` and `teffects psmatch`.

Quick start

Box plot of the propensity score from the last `teffects psmatch` command

```
tebalance box
```

Box plot of values of `x1` in the treatment and control groups from raw data and the matched sample from the last `teffects nnmatch` or `teffects psmatch` command

```
tebalance box x1
```

Menu

Statistics > Causal inference/treatment effects > Balance > Graphs

Syntax

Box plots for the propensity score

```
tebalance box [ , options ]
```

Box plots for a covariate

```
tebalance box varname [ , options ]
```

options

Description

Main

boxlook_options

graph box options controlling how the box looks

legending_options

graph box options controlling how the variables are labeled

axis_options

graph box options controlling how numerical *y* axis is labeled

title_and_other_options

graph box options controlling titles, added text, aspect ratio, etc.

by_options

suboptions inside `by()` controlling plots by raw and matched samples

Options

Main

boxlook_options are any of the options documented in *boxlook_options* in [G-2] **graph box**.

legending_options are any of the options documented in *legending_options* in [G-2] **graph box**.

axis_options are any of the options documented in *axis_options* in [G-2] **graph box**.

title_and_other_options are any of the options, except `by()`, documented in *title_and_other_options* in [G-2] **graph box**. `tebalance box` uses `by()` to differentiate between raw and matched samples, and some *twoway_options* will be repeated for `by graph` and might be better specified as `byopts()`.

by_options are any of the *byopts* documented in [G-3] *by_option*. `byopts()` generally affects the entire graph, and some *by_options* may be better specified as *twoway_options*; see [G-3] **twoway_options**.

Remarks and examples

[stata.com](http://www.stata.com)

When the distribution of a covariate does not vary over the treatment levels, the covariate is said to be balanced. `tebalance box` produces box plots of a covariate over treatment levels for the raw data and for the matched sample produced by `teffects`. If the matched-sample box plots are the same over the treatment levels, the covariate is balanced in the matched sample.

After `teffects nnmatch` and `teffects psmatch`,

```
. tebalance box varname [ , options ]
```

produces box plots to check whether *varname* is balanced.

After `teffects psmatch`,

```
. tebalance box [, options]
```

produces box plots to check whether the propensity score estimated by `teffects` is balanced.

We recommend that you read [CAUSAL] [tebalance](#) before proceeding; it provides an introduction to covariate balance and an overview of the implemented methods.

▷ Example 1: Checking covariate balance after `psmatch`

Using an extract from the data used by [Cattaneo \(2010\)](#), we use `teffects psmatch` to estimate the effect of a mother's smoking behavior (`mbsmoke`) on the birthweight of her child (`bweight`), controlling for marital status (`mmarried`), the mother's age (`mage`), whether the mother had a prenatal doctor's visit in the baby's first trimester (`prenatal1`), and whether this baby is the mother's first child (`fbaby`).

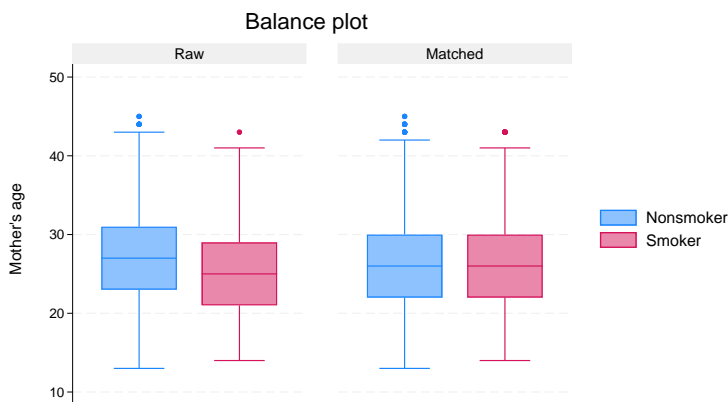
```
. use https://www.stata-press.com/data/r18/cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects psmatch (bweight) (mbsmoke mmarried mage prenatal1 fbaby),
> generate(matchv)

Treatment-effects estimation      Number of obs      =      4,642
Estimator      : propensity-score matching      Matches: requested =      1
Outcome model  : matching                      min =      1
Treatment model: logit                        max =      139
```

bweight		AI robust		z	P> z	[95% conf. interval]
	Coefficient	std. err.				
ATE						
mbsmoke (Smoker vs Nonsmoker)	-235.1714	27.74409	-8.48	0.000	-289.5488	-180.794

We specified the option `generate(matchv)` to speed up the postestimation command that produces density plots, as discussed in [example 3](#) under *Remarks and examples* of [CAUSAL] [tebalance](#). We do not interpret the estimated effect produced by this preliminary model but rather check the specification. Now we look at the box plots.

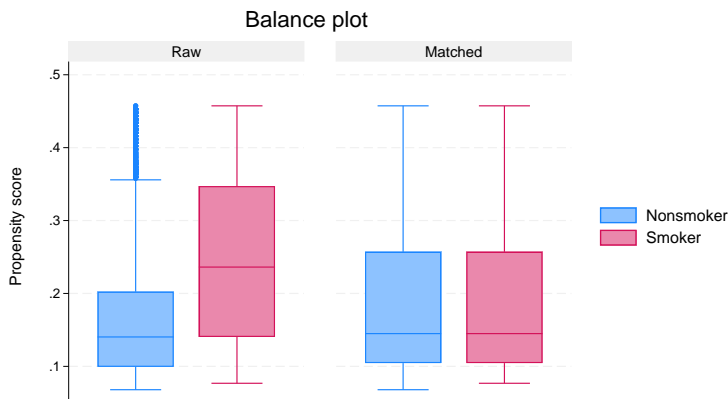
```
. tebalance box mage
```



The box plots for the matched sample are very similar. The medians, the 25th percentiles, and the 75th percentiles appear to be the same, although there may be some differences in the tails, the upper adjacent values, the lower adjacent values, and the outliers. Matching on the estimated propensity score appears to have balanced `mage`, except for the tails.

To get an idea of whether `teffects psmatch` balanced all the covariates, we look at the box plots for the estimated propensity score.

```
. tebalance box
```



The box plots indicate that `teffects psmatch` balanced the estimated propensity scores.

4

Reference

Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154. <https://doi.org/10.1016/j.jeconom.2009.09.023>.

Also see

[CAUSAL] **tebalance** — Check balance after teffects or stteffects estimation

[CAUSAL] **teffects nnmatch** — Nearest-neighbor matching

[CAUSAL] **teffects psmatch** — Propensity-score matching

[CAUSAL] **teoverlap** — Overlap plots

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

