

spbalance — Make panel data strongly balanced

[Description](#)[Quick start](#)[Menu](#)[Syntax](#)[Remarks and examples](#)[Stored results](#)[Also see](#)

Description

`spbalance` reports whether panel data are strongly balanced and, optionally, makes them balanced if they are not.

The data are required to be `xtset`.

Quick start

Determine whether data are strongly balanced

```
spbalance
```

Make data strongly balanced

```
spbalance, balance
```

Menu

Statistics > Spatial autoregressive models

Syntax

Query whether data are strongly balanced

```
spbalance
```

Make data strongly balanced if they are not

```
spbalance, balance
```

Remarks and examples

[stata.com](#)

`Sp` works with panel data but requires that they be strongly balanced. Panels are strongly balanced when each has the same number of observations and defines the same set of times. You can use `spbalance` before data are `spset` or after. Setting the data after is important because `Sp` data that were balanced can become unbalanced after merging additional data.

The data must be `xtset` before you can use `spbalance`:

```
. use http://www.stata-press.com/data/r15/counties
. spbalance
data not xtset
r(459);
. xtset fips time
    panel variable:  fips (unbalanced)
    time variable:  time, 1 to 5, but with a gap
                   delta: 1 unit
. spbalance
(data not strongly balanced)
Type spbalance, balance to make the data strongly balanced by dropping
observations.
```

Type `spbalance`, `balance` to make the data strongly balanced by dropping observations.

```
. spbalance, balance
balancing data ...
2,999 observations dropped.  Dropped was time == 3.  Data are now
strongly balanced.
```

The dataset we started with contained data on five time periods for more than 3,000 U.S. counties. Evidently, some of the panels did not have an observation for time 3. Now, none of the panels have data on time 3. If some panels had no observations on time 4, then all observations for time 4 would have been dropped too.

Balancing by dropping spatial units

`spbalance` balances data by dropping observations for time periods that do not appear in all panels. `spbalance` does not consider the alternative of balancing by dropping spatial units, but you may want to. Here's an example.

We [downloaded shapefiles](#) for all U.S. counties in 2010. We use `spshape2dta` to create Stata Sp datasets:

```
. spshape2dta County_2010Census_DP1
(importing .shp file)
(importing .dbf file)
(creating _ID spatial-unit id)
(creating _CX coordinate)
(creating _CY coordinate)

file County_2010Census_DP1_shp.dta created
file County_2010Census_DP1.dta      created
```

Our analysis dataset is `cbp05_14co.dta` consisting of U.S. Census County Business Patterns data for the years 2005–2014. We load this dataset and merge into it the Sp dataset `County_2010Census_DP1.dta` created by `spshape2dta`.

```
. copy http://www.stata-press.com/data/r15/cbp05_14co.dta .
. use cbp05_14co, clear
```

```
. merge m:1 GEOID10 using County_2010Census_DP1
      Result                # of obs.
-----
not matched                444
  from master              327  (_merge==1)
  from using               117  (_merge==2)
matched                    31,035  (_merge==3)
```

```
. keep if _merge == 3
(444 observations deleted)
. drop _merge
. save cbp05_14co_census
file cbp05_14co_census.dta saved
```

We `xtset` the data and check to see if it is balanced.

```
. xtset _ID year
      panel variable:  _ID (unbalanced)
      time variable:  year, 2005 to 2014
                  delta: 1 unit

. spbalance
(data not strongly balanced)
Type spbalance, balance to make the data strongly balanced by dropping
observations.
```

Both `xtset` and `spbalance` tell us the same thing: the data are unbalanced. We use `spbalance`, `balance` to balance it.

```
. spbalance, balance
balancing data ...
  15,515 observations dropped. Dropped were year == 2005, 2006, 2007,
  2008, 2009. Data are now strongly balanced.
```

What? It dropped all the years 2005–2009.

Let's go back and see what was causing the data to be unbalanced.

```
. use cbp05_14co_census, clear
. bysort _ID: gen npanel = _N
. tabulate npanel
```

npanel	Freq.	Percent	Cum.
5	5	0.02	0.02
10	31,030	99.98	100.00
Total	31,035	100.00	

Every value of `_ID` has data for 10 years except one. The one exception has data for only 5 years. We list it.

```
. list _ID state countyname year npanel if npanel != 10, noobs
```

_ID	state	countyname	year	npanel
400	ND	Slope County	2010	5
400	ND	Slope County	2011	5
400	ND	Slope County	2012	5
400	ND	Slope County	2013	5
400	ND	Slope County	2014	5

Evidently, in the 2010 Census, North Dakota got a new county named Slope County. If we drop it, our data will be balanced.

```
. drop if _ID == 400
(5 observations deleted)

. xtset _ID year
      panel variable:  _ID (strongly balanced)
      time variable:  year, 2005 to 2014
                   delta: 1 unit

. sbalance
(data strongly balanced)
```

There are consequences to this. We dropped a county in the years 2010–2014, and now there is a “hole” in the spatial map for 2010–2014. The county we dropped was part of a larger county before 2010. The spatial maps for this part of North Dakota do not match pre- and post-2010. We might not care about it and just go ahead with our analysis. Or, we might do more work to match up the spatial maps.

This is why `spbalance` always drops times. When it does that, the spatial maps are always the same for the remaining times.

Stored results

`spbalance` without the `balance` option stores the following in `r()`:

Scalars

`r(balanced)` 1 if strongly balanced, 0 otherwise

`spbalance, balance` stores the following in `r()`:

Scalars

`r(balanced)` 1
`r(Ndropped)` number of observations dropped

Matrices

`r(T)` $1 \times r(Ndropped)$ vector of the times dropped if `r(Ndropped) > 0`

Also see

[SP] [intro](#) — Introduction to spatial data and SAR models

[SP] [spset](#) — Declare data to be Sp spatial data

[SP] [spregress](#) — Spatial autoregressive models

[SP] [spxtregress](#) — Spatial autoregressive models for panel data

[XT] [xtset](#) — Declare data to be panel data