Intro 5 — Preparing data: Data containing locations (no shapefiles)

Description Remarks and examples Also see

Description

If you have data that already contain the coordinates of the geographical units, you are not required to use the shapefiles discussed in [SP] **Intro 4**. You may, however, want to use shapefiles. Without them, you cannot create contiguity weighting matrices (matrices in which spillovers occur only among adjacent places), nor can you draw choropleth maps.

This entry provides steps to prepare data with no shapefiles for use by Sp.

You may also be interested in introductions to other aspects of Sp. Below, we provide links to those other introductions.

Intro 1	A brief introduction to SAR models
Intro 2	The **W** matrix
Intro 3	Preparing data for analysis
Intro 4	Preparing data: Data with shapefiles
Intro 6	Preparing data: Data without shapefiles or locations
Intro 7	Example from start to finish
Intro 8	The Sp estimation commands

Remarks and examples

Remarks are presented under the following headings:

Preparation of cross-sectional data
Preparation of panel data
There are no rules as there are with shapefiles

Preparation of cross-sectional data

We will assume that you have file `project_cs2.dta`, which is a cross-sectional dataset on US counties over time, variable `fips` containing the standard county codes, and variables `locx` and `locy` identifying the location of each county.

To turn `project_cs2.dta` into Sp data, do the following:

Step 1: Load the dataset

```
. use project_cs2, clear
```

Step 2: Verify that `fips` is an ID variable

```
. assert fips!=.
. bysort fips: assert _N==1
```

Step 3: `spset` the data

```
. spset fips, coord(locx locy)
```

Step 4: Set the coordinate units, if necessary

```
. spset, coordsys(latlong, miles)
```

Step 5: Save the data

```
. save, replace
```

That is all there is to it.

In step 3, we specified option `coord(locx locy)`. `spset` will create new variables `_ID`, `_CX`, and `_CY`. It will copy `fips` into `_ID`, and `locx` and `locy` into `_CX` and `_CY`.

In step 4, we set the coordinate system to degrees latitude and longitude because that was necessary in this case. We discussed in [SP] **Intro 4** how to determine the coordinate system.

In step 5, we saved `project_cs2.dta` over itself. The new dataset differs from the old in that it has three new variables and is `spset`. No changes or deletions were made to the data.

Preparation of panel data

This time, suppose `project_panel2.dta` is a panel dataset on US counties over time. Perhaps it is already `xtset` on `fips` and `time`. The dataset also includes variables `locx` and `locy` identifying the location of each county.

To turn `project_panel2.dta` into Sp data, do the following:

Step 1: Load the dataset

```
. use project_panel2, clear
```

Step 2: Verify that `fips` and `time` jointly identify the observations

```
. assert fips!=.
. assert time!=.
. bysort fips time: assert _N==1
```

Step 2a: `xtset` the data and verify that `locx` and `locy` are constant within panel

```
. xtset, clear
. xtset fips time
. bysort fips (time): assert locx == locx[1]
. bysort fips (time): assert locy == locy[1]
```

Step 3: Balance and `spset` the data

```
. spbalance
. spset fips, coord(locx locy)
```

Step 4: Set the coordinate units, if necessary

```
. spset, coordsys(latlong)   // optional
```

Step 5: Save the data

```
. save, replace    or    save newfilename, replace
```

Concerning step 5, type `save, replace` only if step 3 did not involve dropping data.

In step 3, we `spset` the data, but not before verifying that they are strongly balanced. If the data are not strongly balanced, `spbalance` will issue an error and suggest that you type

```
. spbalance, balance
```

If you type that, `spbalance` will balance the data.

Then we `spset` the data. This creates the new variables `_ID`, `_CX`, and `_CY`. `spset` copies `fips` into `_ID` and copies `locx` and `locy` into `_CX` and `_CY`.

In step 4, we set coordinate units to degrees latitude and longitude. We discussed how to determine coordinate units in [SP] **Intro 4**.

There are no rules as there are with shapefiles

There are no special rules for working with the data created here as there were when working with data and shapefiles. The rules in [SP] **Intro 4** arose because of the linkage between the data file and its linked `*_shp.dta` file.

Also see

[SP] **spbalance** — Make panel data strongly balanced

[SP] **spset** — Declare data to be Sp spatial data

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.