Title

> **spgenerate —** Generate new variables containing spatial lags

Description

spgenerate creates new variables containing $\mathbf{Wx}$. These are the same spatial lag variables that you include in models that you fit with the Sp estimation commands.

Quick start

Create variable x_nearby equal to Wc*x, the spatial lag of x using spatial weighting matrix Wc

```
spgenerate x_nearby = Wc*x
```

Menu

Statistics > Spatial autoregressive models

Syntax

> spgenerate [*type*] *newvar* = *spmatname**varname [*if*] [*in*]

Remarks and examples

Remarks are presented under the following headings:

> Use with Sp data
> Use with other datasets

Use with Sp data

The $\mathbf{Wx}$ variables that spgenerate creates are literally the variables that the Sp estimation commands include in the models when $\mathbf{x}$ is not the dependent variable. Nonetheless, do not type

```
. spmatrix create contiguity W
. spgenerate Wcollege = W*college
. spregress unemployment college Wcollege, gs2sls
```

Instead, type

```
. spmatrix create contiguity W
. spregress unemployment college, gs2sls ivarlag(W:college)
```

spregress will report the same result either way because college is an exogenous variable. But some postestimation commands will produce incorrect results because they will not know that Wcollege is W*college.

You can use Wcollege after fitting models, however, to better understand results.

In an example in *Fitting models with a spatial lag of independent variables* of [SP] **intro 7**, we fit the model

```
. use texas_ue
. spmatrix create contiguity W
. spregress unemployment college, gs2sls ivarlag(W:college)
  (254 observations)
  (254 observations (places) used)
  (weighting matrix defines 254 places)
```

Spatial autoregressive model	Number of obs	=	254
GS2SLS estimates	Wald chi2(2)	=	81.13
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.2421

unemployment	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
unemployment						
college	-.077997	.0138127	-5.65	0.000	-.1050695	-.0509245
_cons	7.424453	.3212299	23.11	0.000	6.794854	8.054053
W						
college	-.0823959	.0191586	-4.30	0.000	-.1199461	-.0448458

Wald test of spatial terms: chi2(1) = 18.50   Prob > chi2 = 0.0000

Matrix W is the contiguity matrix for first-order neighbors.

If W*college is something of a mystery to you, you can use spgenerate to create the variable and explore it. Type

```
. spgenerate Wcollege = W*college
```

In this example, variables college and Wcollege have similar summary statistics. They usually do.

```
. summarize unemployment college Wcollege
```

Variable	Obs	Mean	Std. Dev.	Min	Max
unemployment	254	4.731102	1.716514	1.5	12.4
college	254	17.95906	7.355919	2.6	49.4
Wcollege	254	15.68765	5.303385	1.279117	36.43961

It turns out that variables college and Wcollege have a surprisingly low correlation, which is not typical:

```
. correlate unemployment college Wcollege
(obs=254)
```

	unempl~t	college	Wcollege
unemployment	1.0000		
college	-0.4323	1.0000	
Wcollege	-0.3833	0.3852	1.0000

You can use Wcollege to assess practical significance. We know from the regression output that the coefficient on W*college is -0.0824 and statistically significant. Is -0.0824 practically significant? From the summarize output, we know that the mean of Wcollege is 15.69. Thus at its average, W*college is contributing $-0.0824 \times 15.69 = -1.29$ to unemployment, which itself has mean 4.73.

Use with other datasets

Consider another analysis that has nothing to do with the spatial analyses discussed in this manual. You are fitting a logistic regression model using outcome.dta. The dataset contains observations on thousands of people whom you call subjects. It has lots of variables, too, among which is fips, the county code in which each subject resides. You want to include the county unemployment rate as an exogenous variable in your model, but outcome.dta does not have that variable.

Obtaining unemployment would be easy enough if you had another dataset containing it, and you do. You have ue_texas.dta, the Sp dataset you used to fit the spatial model above. It is irrelevant that the dataset is spatial; you just want to borrow its county unemployment variable. You could type

```
. use texas_ue, clear
. keep fips unemployment
. save unemploymentvar
. use outcome, clear
. sort fips
. merge m:1 fips using unemploymentvar, keep(master)
. erase unemploymentvar.dta
. logistic outcome ... unemployment ...
```

You had to perform an m:1 merge because outcome.dta might contain multiple subjects living in the same county. You had to keep(master) because there might be some counties in which no one in the data lived. None of that bothers you—you just want the unemployment for the county in which each subject resides, and now you have it, and you fit your model.

What you may not know is that you can include spatial lags of unemployment as an exogenous variable in your logistic model and be on firm statistical ground. A spatial lag is W*unemployment, and W is fixed and unemployment is exogenous in your logistic model. To do that, you would type

```
. use texas_ue, clear
. spmatrix create contiguity W
. spgenerate Wunemployment = W*unemployment
. keep fips unemployment Wunemployment
. save unemploymentvar
. use outcome, clear
. sort fips
. merge m:1 fips using unemploymentvar, keep(master)
. erase unemploymentvar.dta
. logistic outcome ... unemployment Wunemployment ...
```

Also see