Title

stata.com

separate — Create separate variables

Description

`separate` creates new variables containing values from *varname*.

Quick start

Create one variable for each level of `catvar` containing value of `v1` or missing

```
separate v1, by(catvar)
```

As above, but treat missing values of `catvar` as a valid category

```
separate v1, by(catvar) missing
```

Create `v10` as the value of `v1` when `v2` ≥ 20 or missing and missing otherwise and `v11` as the value of `v1` when `v2` < 20 and missing otherwise

```
separate v1, by(v2 < 20)
```

As above, but name new variables `newv1` and `newv2`

```
separate v1, by(v2 < 20) generate(newv) sequential
```

Menu

Data > Create or change data > Other variable-transformation commands > Create separate variables

Syntax

```
separate varname [if] [in], by(groupvar | exp) [options]
```

options	Description
Main	
* `by(`*groupvar*`)`	categorize observations into groups defined by *groupvar*
* `by(`*exp*`)`	categorize observations into two groups defined by *exp*
Options	
`generate(`*stubname*`)`	name new variables by suffixing values to *stubname*; default is to use *varname* as prefix
`sequential`	use as name suffix categories numbered sequentially from 1
`missing`	create variables for the missing values
`shortlabel`	create shorter variable labels

* Either `by(`*groupvar*`)` or `by(`*exp*`)` must be specified.

Options

Main

`by(`*groupvar* | *exp*`)` specifies one variable defining the categories or a logical expression that categorizes the observations into two groups.

If `by(`*groupvar*`)` is specified, *groupvar* may be a numeric or string variable taking on any values.

If `by(`*exp*`)` is specified, the expression must evaluate to true (1), false (0), or missing.

`by()` is required.

Options

`generate(`*stubname*`)` specifies how the new variables are to be named. If `generate()` is not specified, `separate` uses the name of the original variable, shortening it if necessary. If `generate()` is specified, `separate` uses *stubname*. If any of the resulting names is too long when the values are suffixed, it is not shortened and an error message is issued.

`sequential` specifies that categories be numbered sequentially from 1. By default, `separate` uses the actual values recorded in the original variable, if possible, and sequential numbers otherwise. `separate` can use the original values if they are all nonnegative integers smaller than 10,000.

`missing` also creates a variable for the category *missing* if missing occurs (*groupvar* takes on the value missing or *exp* evaluates to missing). The resulting variable is named in the usual manner but with an appended underscore, for example, `bp_`. By default, `separate` creates no such variable. The contents of the other variables are unaffected by whether `missing` is specified.

`shortlabel` creates a variable label that is shorter than the default. By default, when `separate` generates the new variable labels, it includes the name of the variable being separated. `shortlabel` specifies that the variable name be omitted from the new variable labels.

Remarks and examples

stata.com

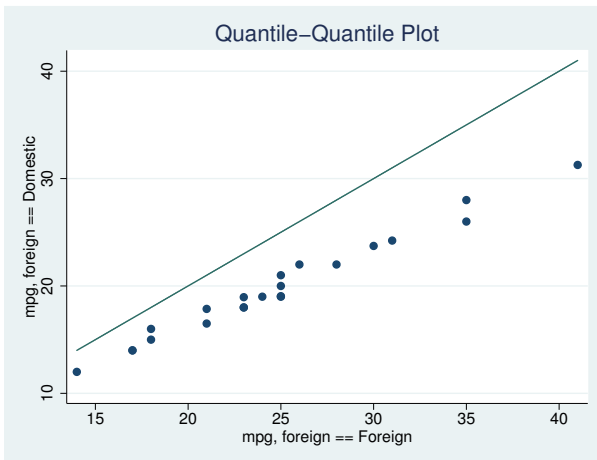
Example 1

We have data on the miles per gallon (mpg) and country of manufacture of 74 automobiles. We want to compare the distributions of mpg for domestic and foreign automobiles by plotting the quantiles of the two distributions (see [R] **diagnostic plots**).

```
. use http://www.stata-press.com/data/r14/auto
(1978 Automobile Data)
. separate mpg, by(foreign)
                storage   display    value
variable name   type      format     label      variable label
------------------------------------------------------------------------------
mpg0            byte      %8.0g                 mpg, foreign == Domestic
mpg1            byte      %8.0g                 mpg, foreign == Foreign
. list mpg* foreign
```

	mpg	mpg0	mpg1	foreign
1.	22	22	.	Domestic
2.	17	17	.	Domestic
3.	22	22	.	Domestic
	(*output omitted*)			
22.	16	16	.	Domestic
23.	17	17	.	Domestic
24.	28	28	.	Domestic
	(*output omitted*)			
73.	25	.	25	Foreign
74.	17	.	17	Foreign

```
. qqplot mpg0 mpg1
```



In our auto dataset, the foreign cars have better gas mileage.

Stored results

`separate` stores the following in `r()`:

Macros
`r(varlist)` names of the newly created variables

Acknowledgment

`separate` was originally written by Nicholas J. Cox of the Department of Geography at Durham University, UK, and coeditor of the *Stata Journal* and author of *Speaking Stata Graphics*.

Reference

Baum, C. F. 2016. *An Introduction to Stata Programming*. 2nd ed. College Station, TX: Stata Press.

Also see

[R] **tabulate oneway** — One-way table of frequencies

[R] **tabulate twoway** — Two-way table of frequencies

[R] **tabulate, summarize()** — One- and two-way tables of summary statistics