

separate — Create separate variables

[Syntax](#)
[Remarks and examples](#)
[Also see](#)

[Menu](#)
[Stored results](#)

[Description](#)
[Acknowledgment](#)

[Options](#)
[Reference](#)

Syntax

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

<i>options</i>	Description
Main	
* by (<i>groupvar</i>)	categorize observations into groups defined by <i>groupvar</i>
* by (<i>exp</i>)	categorize observations into two groups defined by <i>exp</i>
Options	
generate (<i>stubname</i>)	name new variables by suffixing values to <i>stubname</i> ; default is to use <i>varname</i> 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.

Menu

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

Description

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

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](http://www.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/r13/auto
(1978 Automobile Data)
```

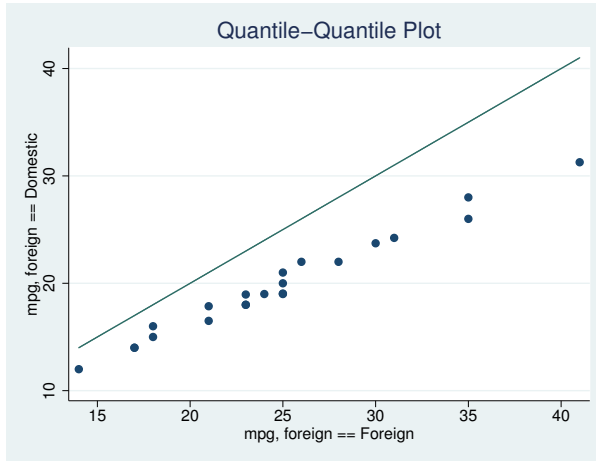
```
. separate mpg, by(foreign)
```

variable name	storage type	display format	value label	variable label
<code>mpg0</code>	byte	<code>%8.0g</code>		<code>mpg, foreign == Domestic</code>
<code>mpg1</code>	byte	<code>%8.0g</code>		<code>mpg, foreign == Foreign</code>

```
. list mpg* foreign
```

	<code>mpg</code>	<code>mpg0</code>	<code>mpg1</code>	<code>foreign</code>
1.	22	22	.	Domestic
2.	17	17	.	Domestic
3.	22	22	.	Domestic
<i>(output omitted)</i>				
22.	16	16	.	Domestic
23.	17	17	.	Domestic
24.	28	28	.	Domestic
<i>(output omitted)</i>				
73.	25	.	25	Foreign
74.	17	.	17	Foreign

```
. qqplot mpg0 mpg1
```



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

4

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*.

Reference

Baum, C. F. 2009. *An Introduction to Stata Programming*. 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