

**separate** — Create separate variables[Description](#)[Options](#)[Reference](#)[Quick start](#)[Remarks and examples](#)[Also see](#)[Menu](#)[Stored results](#)[Syntax](#)[Acknowledgment](#)

## 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)
```

Same 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`  $\geq$  20 or missing and missing otherwise and `v11` as the value of `v1` when `v2`  $<$  20 and missing otherwise

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

Same 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]
```

<i>options</i>	Description
<b>Main</b>	
* <b>by</b> ( <i>groupvar</i> )	categorize observations into groups defined by <i>groupvar</i>
* <b>by</b> ( <i>exp</i> )	categorize observations into two groups defined by <i>exp</i>
<b>Options</b>	
<b>generate</b> ( <i>stubname</i> )	name new variables by suffixing values to <i>stubname</i> ; default is to use <i>varname</i> as prefix
<b>sequential</b>	use as name suffix categories numbered sequentially from 1
<b>missing</b>	create variables for the missing values
<b>shortlabel</b>	create shorter variable labels
* Either <b>by</b> ( <i>groupvar</i> ) or <b>by</b> ( <i>exp</i> ) must be specified.	
collect is allowed; see [U] 11.1.10 <b>Prefix commands</b> .	

## 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

## ▷ 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](#)).

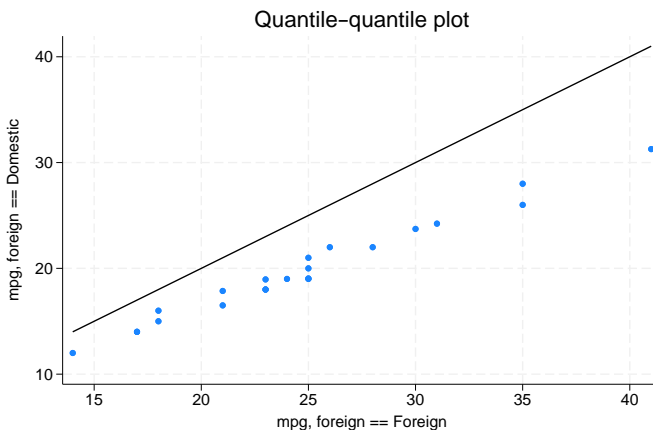
```
. use https://www.stata-press.com/data/r18/auto
(1978 automobile data)
. separate mpg, by(foreign)
```

Variable name	Storage type	Display format	Value 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
<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.

## 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, who is 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

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).