

Stata tip 61: Decimal commas in results output and data input

Nicholas J. Cox
Department of Geography
Durham University
Durham City, UK
n.j.cox@durham.ac.uk

Given a decimal fraction to evaluate, such as $5/4$, Stata by default uses a period (stop) as a decimal separator and shows the result as 1.25. That is, the period separates, and also joins, the integer part 1 and the fractional part 25, meaning here 25/100. Many Stata users, particularly in the United States and several other English-speaking countries, will have learned of such decimal points at an early age and so may think little of this. However, in many other countries, commas are used as decimal separators, so that 1,25 is the preferred way of representing such fractions. This tip is for those users, although it also provides an example of how different user preferences can be accommodated by Stata.

Over several centuries, mathematicians and others have been using decimal fractions without uniformity in their representation. Cajori (1928, 314–335) gave one detailed historical discussion that is nevertheless incomplete. Periods and commas have been the most commonly used separators since the development of printing. Some authors, perhaps most notably John Napier of logarithm fame, even used both symbols in their work. An objection to the period is its common use to indicate multiplication, so some people have preferred centered or even raised dots, even though the same objection can be made in reverse, at least to centered dots. An objection to the comma is similarly to its common use to separate numbers in lists, although serious ambiguity should not arise so long as such lists are suitably spaced out.

Incidentally, many notations other than periods and commas have been proposed and several remained in use until well into the twentieth century. Indeed the momayyez, a mark like a forward slash or comma, is widely used at present in several Middle Eastern countries.

Stata 7 introduced `set dp comma` as a way to set the decimal point to a comma in output. Thus after

```
. set dp comma  
. display 5/4
```

shows 1,25 and other output follows the same rule. Type `set dp comma`, permanently to have such output permanently. Type `set dp period` to restore the default.

Stata 7 also introduced comma-based formats such as `%7,2f`. See the help or manual entry for `format` for more details.

This still leaves the large question of input. Suppose, for example, that you have text or other data files in which numeric variables are indicated with commas as separators. Stata will not accept such variables as numeric, but the solution now is simple. Read in such variables as string, and then within Stata use **destring**, **replace dpcomma**. **destring** is especially convenient because it can be applied to several variables at once.

The **dpcomma** option was added to the **destring** command on 15 October 2007 and so is not documented in the Stata 10 manuals and not implemented in versions earlier than 10. Users still using Stata 9 or earlier are advised to use the **subinstr()** function to change commas to periods, followed by **destring**. Naturally, great care is required if any periods also appear as separators before or after the decimal comma. Any such periods should be removed before the comma is converted to a period.

Reference

Cajori, F. 1928. *A History of Mathematical Notations. Volume I: Notation in Elementary Mathematics*. Chicago: Open Court.