

Working With Dates in Stata

Andy Grogan-Kaylor

7 Apr 2021 12:22:32

Introduction

Dates in any statistical software (Stata, R) are complicated.

For example, a particular date could be coded as “4-5-2021”, or “5-4-2021” or “April 5, 2021” or “5APR2021”.

In addition to the multiplicity of possible formats it is also difficult to do calculations on these kinds of quantities e.g. “how many days have elapsed from Day A to Day B?”

To address these issues, Stata wants these dates to be encoded as a number, specifically the number of days since January 1, 1960. We then make sure to format these numbers as dates.

Get The Data

```
. use "simulated-date.dta", clear
```

List And Describe The Data

We see that both date variables are formatted as strings

```
. list
```

	startdate	enddate
1.	2019-01-01	2019-1-30
2.	2019-02-15	2019-5-30
3.	2019-03-01	2019-4-30

```
. describe
```

Contains data from simulated-date.dta

```
obs:      3
vars:      2                               6 Apr 2021 16:46
```

variable name	storage type	display format	value label	variable label
startdate	str10	%10s		startdate
enddate	str9	%9s		enddate

Sorted by:

Create Date Variables

```
. generate startdate2 = date(startdate, "YMD") // create a date specifying  
> order of elements  
  
. generate enddate2 = date(enddate, "YMD") // create a date specifying ord  
> er of elements
```

The command has created 2 dates in numeric form, but they display as numbers.

```
. describe  
Contains data from simulated-date.dta  
obs:          3  
vars:          4                      6 Apr 2021 16:46
```

variable name	storage type	display format	value label	variable label
startdate	str10	%10s		startdate
enddate	str9	%9s		enddate
startdate2	float	%9.0g		
enddate2	float	%9.0g		

Sorted by:
Note: Dataset has changed since last saved.

```
. list
```

	startdate	enddate	startd_2	enddate2
1.	2019-01-01	2019-1-30	21550	21579
2.	2019-02-15	2019-5-30	21595	21699
3.	2019-03-01	2019-4-30	21609	21669

Format As Dates

```
. format %d startdate2 enddate2  
  
. describe  
Contains data from simulated-date.dta  
obs:          3  
vars:          4                      6 Apr 2021 16:46
```

variable name	storage type	display format	value label	variable label
startdate	str10	%10s		startdate
enddate	str9	%9s		enddate
startdate2	float	%d		
enddate2	float	%d		

Sorted by:
Note: Dataset has changed since last saved.

```
. list
```

	startdate	enddate	startda_2	enddate2
1.	2019-01-01	2019-1-30	01jan2019	30jan2019
2.	2019-02-15	2019-5-30	15feb2019	30may2019
3.	2019-03-01	2019-4-30	01mar2019	30apr2019