Working With Dates in Stata

Andy Grogan-Kaylor

20 Apr 2023 10:16:14

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
 Observations:             3
    Variables:             2                  6 Apr 2021 16:46
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    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 elemen
> ts

. generate enddate2 = date(enddate, "YMD") // create a date specifying order of elements
```

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

```
. describe
Contains data from simulated-date.dta
 Observations:             3
    Variables:             4                  6 Apr 2021 16:46
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    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
 Observations:             3
    Variables:             4                  6 Apr 2021 16:46
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    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

Calculations

We can now use dates in calculations. For example, "How much time has elapsed between `startdate2` and `enddate2`?"

```
. generate elapseddays = enddate2 - startdate2

. generate elapsedyears = (enddate2 - startdate2)/365

. list, abbreviate(15) // list out the data with new variables
```

	startdate	enddate	startdate2	enddate2	elapseddays	elapsedyears
1.	2019-01-01	2019-1-30	01jan2019	30jan2019	29	.0794521
2.	2019-02-15	2019-5-30	15feb2019	30may2019	104	.2849315
3.	2019-03-01	2019-4-30	01mar2019	30apr2019	60	.1643836